

UNIVERSIDADE FEDERAL DO PARANÁ

RODRIGO AMARAL SEGER

MÉTODO PARA DETECÇÃO DE MOVIMENTOS
AUXILIARES DE CLARINETISTAS BASEADO EM
VISÃO COMPUTACIONAL

CURITIBA
2012

RODRIGO AMARAL SEGER

MÉTODO PARA DETECÇÃO DE MOVIMENTOS
AUXILIARES DE CLARINETISTAS BASEADO EM
VISÃO COMPUTACIONAL

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Engenharia Elétrica, Programa de Pós-Graduação em Engenharia Elétrica, Departamento de Engenharia Elétrica, Setor de Tecnologia, Universidade Federal do Paraná.

Orientador: Prof. Dr. Alessandro Lameiras Koerich.

Co-orientador: Prof. Dr. Marcelo Mortensen Wanderley.

Co-orientador: Prof. Dr. Oscar da Costa Gouveia Filho.

CURITIBA
2012

TERMO DE APROVAÇÃO

Rodrigo Amaral Seger

**Método para Detecção de Movimentos Auxiliares de Clarinetistas
Baseado em Visão Computacional**

Dissertação aprovada como requisito parcial para obtenção do grau de Mestre no
Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do
Paraná.

Prof. Dr. Alessandro Lameiras Koerich – Orientador
Universidade Federal do Paraná

Prof. Dr. Alessandro Zimmer – Convidado
Universidade Federal do Paraná

Prof. Dr. Alceu de Souza Britto Jr. – Convidado
Pontifícia Universidade Católica do Paraná

Curitiba, 24 de Abril de 2012

RESUMO

Este trabalho propõe um método para detecção de movimentos auxiliares espontâneos de clarinetistas durante a execução de uma peça musical a partir da análise dos vídeos obtidos durante essas apresentações, utilizando técnicas e algoritmos de visão computacional. Durante a execução de uma música, pode-se observar toda uma gesticulação produzida pelo instrumentista. Além dos movimentos que geram a música, há outros diversos gestos que não são diretamente responsáveis pela produção do som, mas que são encontrados durante o processo. Estes movimentos são chamados gestos auxiliares. Neste trabalho propõe-se uma metodologia baseada apenas na análise de vídeo através de algoritmos de visão computacional, que é capaz de indicar a ocorrência de movimentos auxiliares específicos. O estudo levou à união de técnicas como subtração de fundo, rastreamento de pontos, projeções de perfil e limiarização, que somadas a filtros espaciais e cálculos matemáticos possibilitaram a avaliação de movimentos sobre uma base de dados de nove vídeos com apresentações de clarinetistas em estúdio. O método proposto atingiu cerca de 70% de taxa de acerto na detecção de movimentos de dobra de joelhos, 75% na identificação de movimentos pendulares de clarinete e mais de 80% na avaliação de movimentos das costas.

Palavras-chave: Movimentos de clarinete. Movimentos de dobra de joelhos. Movimentos de ondulação nas costas. Detecção de gestos auxiliares. Técnicas de Visão Computacional.

ABSTRACT

This work aims at developing an automatic method for the detection of clarinetists' ancillary gestures by analyzing videos from musical performances using computer vision techniques and algorithms. Apart from gestures used for the generation of sound in acoustic musical instruments, performers produce several other gestures that are not directly responsible for the production of sound, but that are often found in performances. These gestures have been called ancillary gestures. This research proposes a methodology that is able to indicate the occurrence of these specific gestures, based only on computer vision algorithms. The study led to the combination of techniques such as background subtraction, tracking points, projection profiles and thresholding, which when used with spatial filters and mathematical calculations allowed the evaluation of movements in a database comprised of nine videos. The proposed method reached 70% accuracy in detecting movements of bending knees, 75% in the clarinet bell and more than 80% in the evaluation of back curling gestures.

Keywords: Clarinet bell gestures. Bending knees gestures. Back curling gestures. Ancillary gestures detection. Computer vision techniques.

LISTA DE FIGURAS

Figura 1: Exemplo de movimento de clarinete.	13
Figura 2: Exemplo de movimento de dobra de joelhos.	14
Figura 3: Exemplo de movimento de ondulações nas costas.....	14
Figura 4: Exemplo de deslocamento de peso.	14
Figura 5: Visão geral do método proposto.....	23
Figura 6: Diagrama de blocos da etapa de pré-processamento.....	25
Figura 7: Vídeo original.	26
Figura 8: Subtração de fundo.	27
Figura 9: Região de interesse.	28
Figura 10: Comparação entre vídeo original e vídeo pré-processado.....	29
Figura 11: Rotulação dos vídeos.	31
Figura 12: Diagrama de blocos do método de análise de movimentos de clarinete..	32
Figura 13: Exemplo de deslocamento do marcador virtual rastreado.	34
Figura 14: Trajetória do marcador virtual rastreado: a) componente vertical; b) componente horizontal.	35
Figura 15: Posição do centro de massa ao longo de quadros do vídeo.....	37
Figura 16: Exemplo de curva de diferença sobreposta à rotulação.....	38
Figura 17: Diagrama de blocos do método de análise de movimentos de dobra dos joelhos.....	40
Figura 18: Projeção de perfil horizontal.....	42
Figura 19: Projeção de perfil com dobra de joelhos.	42
Figura 20: Referencial proporcional ao centro de massa.	43
Figura 21: Exemplo de limiarização da projeção de perfil horizontal.....	45
Figura 22: Comparação entre imagem e dados pós-processados.....	46
Figura 23: Diagrama de blocos do método de análise de movimentos de ondulações nas costas.	47
Figura 24: Uso da <i>bounding box</i>	48
Figura 25: Comparação entre <i>bounding box</i> completa e restrita.....	49
Figura 26: Projeção de perfil vertical apenas na região de interesse.....	49
Figura 27: Projeção de perfil vertical apenas na ocorrência de movimento.	50
Figura 28: Pontos de interesse.....	51
Figura 29: Triangulação.....	52
Figura 30: Variação da triangulação durante a presença de movimento.....	53
Figura 31: Diagrama final da metodologia.....	55
Figura 32: Resultado da análise do movimento de clarinete para o vídeo 1.4.	56
Figura 33: Efeito de trem de movimentos.....	60
Figura 34: Resultado da análise do movimento de joelhos para o vídeo 3 (quadros 1 - 343).	62
Figura 35: Resultado da análise do movimento de joelhos para o vídeo 3 (quadros 344 - 686).	63
Figura 36: Resultado da análise do movimento de joelhos para o vídeo 3 (quadros 687 - 1029).	63
Figura 37: Resultado da análise do movimento de joelhos para o vídeo 3 (quadros 1030 - 1372).	63
Figura 38: Resultado da análise do movimento de joelhos para o vídeo 3 (quadros 1373 - 1715).	64
Figura 39: Resultado da análise do movimento de joelhos para o vídeo 3 (quadros 1716 - 2058).	64

Figura 40: Resultado da análise do movimento de ondulação das costas para o vídeo 2.2.....	66
Figura 41: Pulsos alongados no vídeo 4.	68
Figura 42: Atraso de detecção no vídeo 1.1.....	69
Figura 43: Trem de movimentos no vídeo 2.2.	70
Figura 44: Resultado da análise do movimento de clarinete para o vídeo 1.1.	86
Figura 45: Resultado da análise do movimento de clarinete para o vídeo 1.2.	86
Figura 46: Resultado da análise do movimento de clarinete para o vídeo 1.3.	86
Figura 47: Resultado da análise do movimento de clarinete para o vídeo 2.1.	87
Figura 48: Resultado da análise do movimento de clarinete para o vídeo 2.2.	87
Figura 49: Resultado da análise do movimento de clarinete para o vídeo 2.3.	87
Figura 50: Resultado da análise do movimento de clarinete para o vídeo 3.	88
Figura 51: Resultado da análise do movimento de clarinete para o vídeo 4.	88
Figura 52: Resultado da análise do movimento de joelhos para o vídeo 1.3 (quadros 1 - 437).....	91
Figura 53: Resultado da análise do movimento de joelhos para o vídeo 1.3 (quadros 438 - 874).....	91
Figura 54: Resultado da análise do movimento de joelhos para o vídeo 1.3 (quadros 875 - 1311).....	92
Figura 55: Resultado da análise do movimento de joelhos para o vídeo 1.3 (quadros 1749 - 2185).....	92
Figura 56: Resultado da análise do movimento de joelhos para o vídeo 1.3 (quadros 2186 - 2622).....	92
Figura 57: Resultado da análise do movimento de joelhos para o vídeo 2.1 (quadros 1 - 395).....	92
Figura 58: Resultado da análise do movimento de joelhos para o vídeo 2.1 (quadros 396 - 790).....	92
Figura 59: Resultado da análise do movimento de joelhos para o vídeo 2.1 (quadros 791 - 1185).....	93
Figura 60: Resultado da análise do movimento de joelhos para o vídeo 2.1 (quadros 1186 - 1580).....	93
Figura 61: Resultado da análise do movimento de joelhos para o vídeo 2.1 (quadros 1581 - 1975).....	93
Figura 62: Resultado da análise do movimento de joelhos para o vídeo 2.1 (quadros 1976 - 2370).....	93
Figura 63: Resultado da análise do movimento de joelhos para o vídeo 2.2 (quadros 510 - 1072).....	93
Figura 64: Resultado da análise do movimento de joelhos para o vídeo 2.2 (quadros 1073 - 1634).....	94
Figura 65: Resultado da análise do movimento de joelhos para o vídeo 2.2 (quadros 1635 - 2196).....	94
Figura 66: Resultado da análise do movimento de joelhos para o vídeo 2.2 (quadros 2197 - 2758).....	94
Figura 67: Resultado da análise do movimento de joelhos para o vídeo 2.3 (quadros 210 - 646).....	94
Figura 68: Resultado da análise do movimento de joelhos para o vídeo 2.3 (quadros 647 - 1082).....	94
Figura 69: Resultado da análise do movimento de joelhos para o vídeo 2.3 (quadros 1083 - 1518).....	95
Figura 70: Resultado da análise do movimento de joelhos para o vídeo 2.3 (quadros	

1519 - 1954).....	95
Figura 71: Resultado da análise do movimento de joelhos para o vídeo 2.3 (quadros 1955 - 2392).....	95
Figura 72: Resultado da análise do movimento de joelhos para o vídeo 3 (quadros 1 - 337).....	95
Figura 73: Resultado da análise do movimento de joelhos para o vídeo 3 (quadros 338 - 674).....	96
Figura 74: Resultado da análise do movimento de joelhos para o vídeo 3 (quadros 675 - 1011).....	96
Figura 75: Resultado da análise do movimento de joelhos para o vídeo 3 (quadros 1012 - 1348).....	96
Figura 76: Resultado da análise do movimento de joelhos para o vídeo 3 (quadros 1349 - 1685).....	96
Figura 77: Resultado da análise do movimento de ondulação das costas para o vídeo 1.1.....	102
Figura 78: Resultado da análise do movimento de ondulação das costas para o vídeo 1.2.....	102
Figura 79: Resultado da análise do movimento de ondulação das costas para o vídeo 1.3.....	103
Figura 80: Resultado da análise do movimento de ondulação das costas para o vídeo 1.4.....	103
Figura 81: Resultado da análise do movimento de ondulação das costas para o vídeo 2.1.....	103
Figura 82: Resultado da análise do movimento de ondulação das costas para o vídeo 2.3.....	104
Figura 83: Resultado da análise do movimento de ondulação das costas para o vídeo 3.	104
Figura 84: Resultado da análise do movimento de ondulação das costas para o vídeo 4.	104

LISTA DE TABELAS

Tabela 1: Base de vídeos.	25
Tabela 2: Limiares da curva de diferença absoluta para cada vídeo.	39
Tabela 3: Média e desvio padrão da coordenada largura do centro de massa e fator limitador.	44
Tabela 4: Média e desvio padrão do método de triangulação.	54
Tabela 5: Resultado da análise do movimento de clarinete para o vídeo 1.4.	57
Tabela 6: Resultado da análise do movimento de clarinete para o vídeo 3.	59
Tabela 7: Resultado final da análise de movimentos de clarinete.	61
Tabela 8: Resultado da análise do movimento de joelhos para o vídeo 3.	62
Tabela 9: Resultado final da análise de movimentos de dobra de joelhos.	65
Tabela 10: Resultado da análise do movimento de costas para o vídeo 2.2.	67
Tabela 11: Resultado final da análise de movimentos de ondulação das costas.	71
Tabela 12: Resultado da análise do movimento de clarinete para o vídeo 1.1.	79
Tabela 13: Resultado da análise do movimento de clarinete para o vídeo 1.2.	80
Tabela 14: Resultado da análise do movimento de clarinete para o vídeo 1.3.	81
Tabela 15: Resultado da análise do movimento de clarinete para o vídeo 2.1.	82
Tabela 16: Resultado da análise do movimento de clarinete para o vídeo 2.2.	83
Tabela 17: Resultado da análise do movimento de clarinete para o vídeo 2.3.	84
Tabela 18: Resultado da análise do movimento de clarinete para o vídeo 3.	85
Tabela 19: Resultado da análise do movimento de clarinete para o vídeo 4.	85
Tabela 20: Resultado da análise do movimento de joelhos para o vídeo 1.1.	88
Tabela 21: Resultado da análise do movimento de joelhos para o vídeo 2.1.	88
Tabela 22: Resultado da análise do movimento de joelhos para o vídeo 2.2.	89
Tabela 23: Resultado da análise do movimento de joelhos para o vídeo 2.3.	90
Tabela 24: Resultado da análise do movimento de joelhos para o vídeo 4.	91
Tabela 25: Resultado da análise do movimento de ondulação das costas para o vídeo 1.1.	97
Tabela 26: Resultado da análise do movimento de ondulação das costas para o vídeo 1.2.	98
Tabela 27: Resultado da análise do movimento de ondulação das costas para o vídeo 1.3.	99
Tabela 28: Resultado da análise do movimento de ondulação das costas para o vídeo 1.4.	100
Tabela 29: Resultado da análise do movimento de ondulação das costas para o vídeo 2.3.	101
Tabela 30: Resultado da análise do movimento de ondulação das costas para o vídeo 3.	101
Tabela 31: Resultado da análise do movimento de ondulação das costas para o vídeo 4.	102

SUMÁRIO

1 INTRODUÇÃO	11
1.1 MOTIVAÇÃO	11
1.2 DEFINIÇÃO DO PROBLEMA	11
1.3 HIPÓTESE	16
1.4 OBJETIVOS	16
1.4.1 Objetivo Geral	16
1.4.2 Objetivos Específicos	16
1.5 CONTRIBUIÇÕES	17
1.6 ESTRUTURA	17
2 ESTADO DA ARTE	18
3 MÉTODO PROPOSTO	23
3.1 VISÃO GERAL	23
3.2 BASE DE DADOS	24
3.3 PRÉ-PROCESSAMENTO	25
3.4 ROTULAÇÃO	30
3.5 DETECÇÃO DE MOVIMENTOS DE INTERESSE	32
3.5.1 Movimentos de Clarinete	32
3.5.2 Movimentos de Dobra de Joelhos	40
3.5.3 Movimentos de Ondulações nas Costas	46
4 AVALIAÇÃO DOS RESULTADOS	56
4.1 ANÁLISE DE MOVIMENTOS DE CLARINETE	56
4.2 ANÁLISE DE MOVIMENTOS DE DOBRA DE JOELHOS	62
4.3 ANÁLISE DE MOVIMENTOS DE ONDULAÇÕES NAS COSTAS	66
5 CONCLUSÃO	72
REFERÊNCIAS	75
APÊNDICE	79

1 INTRODUÇÃO

O estudo da interação corporal humana com a música é assunto bastante atual e que desperta cada vez mais interesse de pesquisadores de diversas áreas como engenharia, computação, música e psicologia (GUNES; PICCARDI, 2007; JENSENIUS; WANDERLEY; GODOY; LEMAN, 2010; TEIXEIRA, 2010). Trata-se de um assunto complexo e ainda pouco compreendido pela comunidade científica.

1.1 MOTIVAÇÃO

O interesse no assunto advém de dois fatores principais: o primeiro deles deve-se às aplicações emergentes em realidade virtual que requerem sistemas robustos e confiáveis, necessitando constante evolução dos sistemas de aquisição e reconhecimento a fim de tornar a interação mais agradável, interessante e real (MAZZARINO; PEINADO; BOULIC; VOLPE; WANDERLEY, 2009).

O segundo, não menos importante, relaciona-se ao estudo comportamental humano. Por que agimos de determinada maneira quando ouvimos alguma música? Por que nos movimentamos durante a execução de uma peça musical?

As duas justificativas são igualmente importantes e demandam um processo em comum: ferramentas que possibilitem o estudo. É neste contexto que esta pesquisa está inserida: na formulação de uma ferramenta computacional que colabore com estes estudos.

1.2 DEFINIÇÃO DO PROBLEMA

Quando se analisa comportamento humano, um conceito vem à tona: gestos. Basicamente a interação do ser humano com o mundo a seu redor se faz através de gestos (AGGARWAL; CAI, 1999; GAVRILA, 1999; MITRA; ACHARYA, 2007). Seja numa simples conversa, ou em práticas esportivas, em danças, num discurso, ou dirigindo, enfim, tudo é formado pela reunião de diferentes gestos, e numa *performance* musical isto não é diferente. Nota-se claramente que esta é repleta de movimentos.

Centrado no estudo dos gestos presentes na música, pesquisas sugerem que estes se dividem basicamente em três grandes grupos: os gestos que produzem efetivamente som, os que acompanham o processo de produção sonora e os meramente figurativos (DELALANDE, 1988).

Movimentos que produzem som são intuitivos e, até certo ponto, evidentes, o que não é verdade para os movimentos de acompanhamento. Por que muitas vezes agimos de forma não intencional, espontânea, com determinados movimentos em função de uma fala, ou de uma execução musical? Esta é a grande motivação para as pesquisas atuais. É interessante observar como praticamente ninguém discursa ou produz música de forma estática, inerte, a não ser que isso seja feito propositadamente.

Movimentos do tronco, dos cotovelos, mímicas e respirações (DELALANDE, 1988), além de passos e expressões faciais, em geral não atuam diretamente na produção sonora, mas acompanham o processo (isto justifica o termo “gestos de acompanhamento”). Estes movimentos recebem tanta atenção justamente por se tratarem de gestos aparentemente ilógicos e inúteis, que teoricamente não possuem motivo, mas que ocorrem de forma natural e espontânea, sugerindo algum significado, algum auxílio na construção da fala, ou da música. Por esta razão, os gestos de acompanhamento são também conhecidos como gestos não óbvios, ou ainda como gestos auxiliares (WANDERLEY, 1999).

Avaliar todo um universo de músicos e seus movimentos é algo impraticável. Geralmente os trabalhos restringem seu escopo a certa categoria de músicos e/ou peças musicais. Em Delalande 1988 os estudos focaram em peças para piano, já Cadoz e Wanderley 2000 apresentaram estudos com recitais de cello, gaita de fole e clarinete. Já em Wanderley 2002, os movimentos foram observados durante a execução de um trecho da Primeira Sonata para Clarinete Op. 120 de Brahms, mas em Wanderley, et. al. 2005 o estudo de movimentos auxiliares de clarinetistas especificou-se para um trecho de cerca de 1 minuto do Segundo Movimento das Três Peças de Stravinsky para Clarinete Solo em um estúdio. Nove vídeos desta base de 2005 foram cedidos para esta pesquisa e a partir destes buscar-se-á identificar alguns movimentos auxiliares produzidos nas apresentações.

Quando se trata de clarinetistas, diversos movimentos aparecem como inerentes à classe. Por exemplo: movimento completo cima/baixo do clarinete,

movimento circular completo do clarinete, movimento cima/baixo de cabeça, movimento cima/baixo de ombros, movimento de ondulação das costas, movimento de abertura/fechamento de braços (“bater asas”), movimento de inclinação da cintura, movimento de dobra de joelhos, movimento de inclinação e batida de pés, deslocamento de peso (balanço do corpo) para a esquerda ou para a direita, entre outros (WANDERLEY; VINES; MIDDLETON; McKAY; HATCH, 2005). Estes movimentos não produzem, de fato, o som do instrumento, mas são encontrados com frequência em observações e, por isso, são tachados como movimentos auxiliares.

Especificando esta coleção de gestos, quatro deles surgem como característicos desta classe de instrumentistas, e ocorrem com maior frequência: o movimento completo cima/baixo do clarinete, o movimento de dobra de joelhos, o movimento de ondulação das costas e o deslocamento de peso (WANDERLEY, et al., 2005; VERFAILLE; QUEK; WANDERLEY, 2006 ; JENSENIUS, et al., 2010) e por esta razão passaram a ser o foco deste estudo. O movimento de clarinete é uma transição do instrumento musical, conforme ilustra a Figura 1, o movimento de dobra de joelhos é o próprio abaixamento do instrumentista, exemplificado na Figura 2, o movimento de ondulação das costas é a inclinação momentânea do tronco do músico para frente, demonstrado na Figura 3 e o deslocamento de peso é a transferência do equilíbrio corporal de uma perna para a outra, ilustrado na Figura 4.

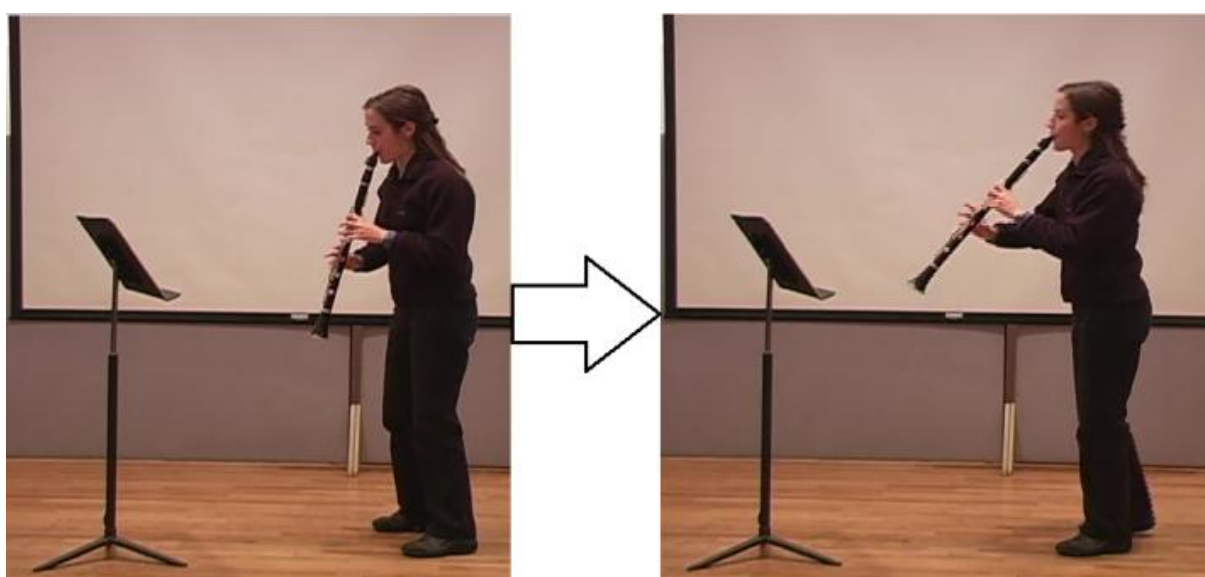


Figura 1: Exemplo de movimento de clarinete.

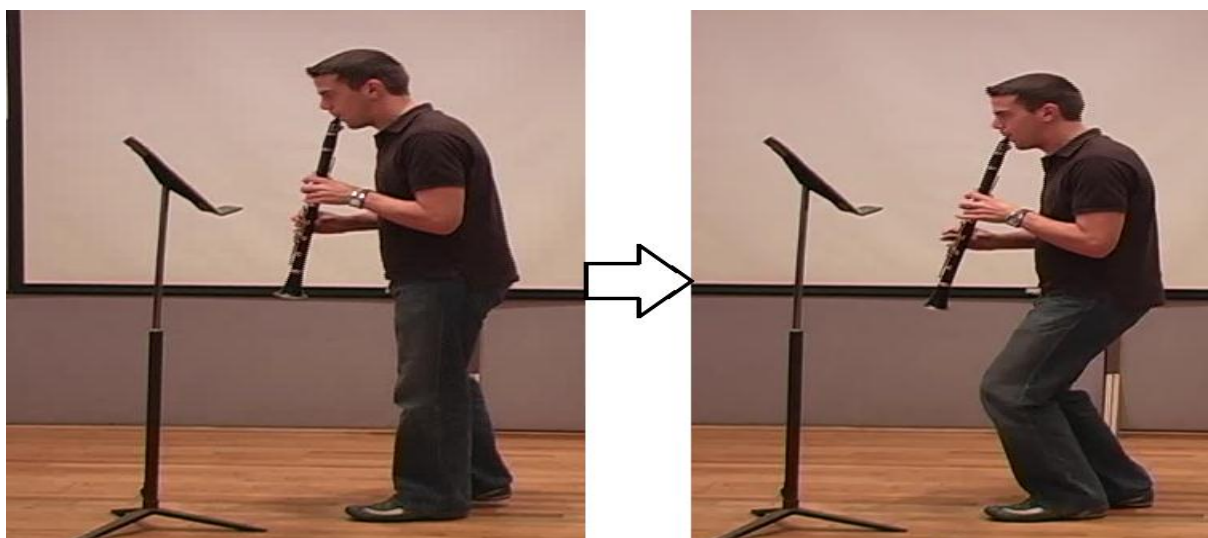


Figura 2: Exemplo de movimento de dobra de joelhos.

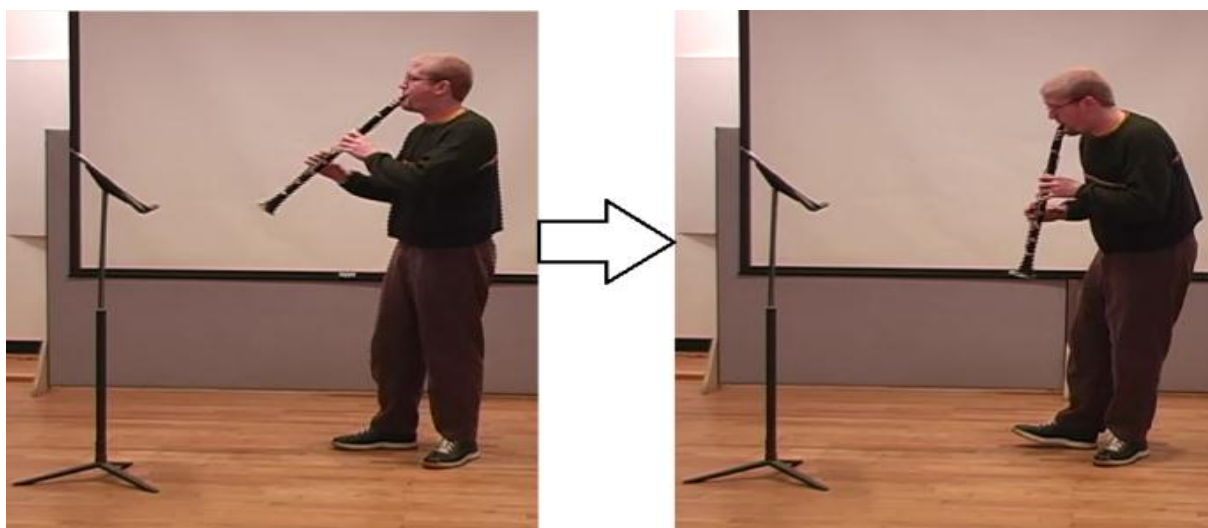


Figura 3: Exemplo de movimento de ondulações nas costas.

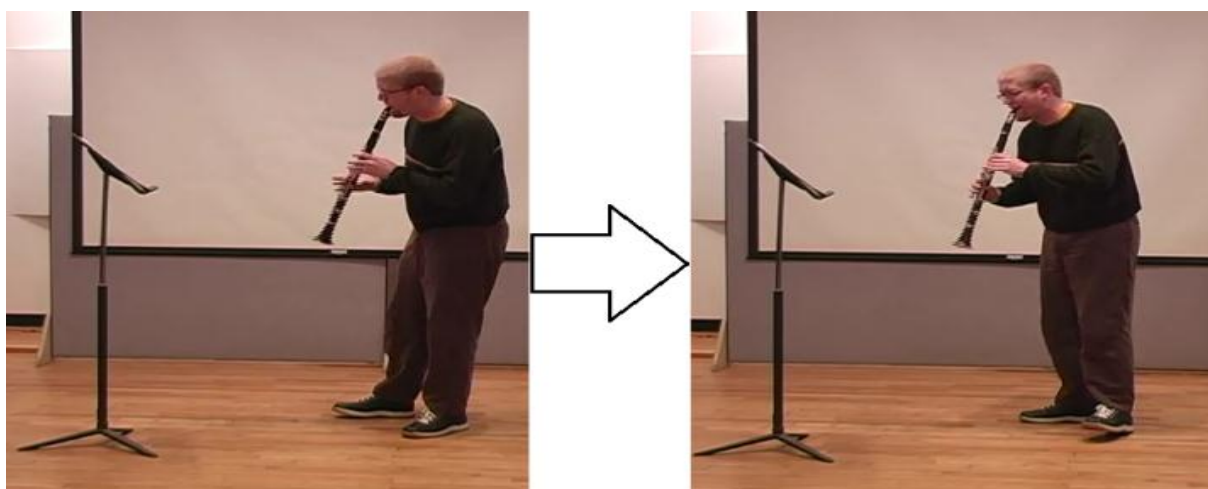


Figura 4: Exemplo de deslocamento de peso.

Atualmente a detecção desses movimentos auxiliares é feita através da avaliação do pesquisador a partir da observação de vídeos e do estudo de dados obtidos via sensores de movimento fixados aos músicos e ao instrumento musical (WANDERLEY, 1999; CADOZ; WANDERLEY, 2000; WANDERLEY, 2002; DAHL; FRIBERG, 2004; WANDERLEY, et al., 2005; VERFAILLE, et al., 2006; JENSENIUS, et al., 2010; TEIXEIRA, 2010).

Observar o gestual produzido durante uma apresentação musical é muito trabalhoso e subjetivo quando feito por pesquisadores assistindo aos vídeos dos músicos. A quantidade de informação é tamanha que se torna estafante detectar e avaliar movimentos apenas reproduzindo repetidamente as gravações, sendo o resultado dependente do avaliador (pessoas diferentes podem caracterizar os movimentos diferentemente). Além do mais, o excessivo número de dados obtidos via sensores, dados estes às vezes até desnecessários, onera o processo de pesquisa, seja em capacidade de processamento ou simplesmente tempo de análise.

Um método capaz de detectar de forma computacional os movimentos auxiliares através de vídeos de apresentações gravados seria uma solução adequada para eliminar a atual necessidade de visualização das gravações e, assim, facilitar as análises pormenorizadas.

Esta solução é viável, contudo, gera fatores complicadores. Como cada apresentação é única, particular, é muito difícil aplicar um algoritmo padrão a todos. Não há como generalizar resultados, uma vez que cada vídeo de clarinetista apresenta características únicas de cenário, iluminação e distanciamento em relação à câmera, além de as pessoas filmadas também serem diferentes umas das outras, com aspectos físicos distintos, roupas diferentes e até mesmo agirem diferentemente.

Além disso, a base de vídeos foi gerada por apenas uma câmera, em um plano de gravação. Assim, dentre os quatro gestos auxiliares característicos listados, o movimento de deslocamento de peso não pode ser estudado, uma vez que a ação ocorre em três dimensões. Como não há informação de profundidade disponível, é inviável a detecção deste movimento. Dessa forma, apenas os movimentos de clarinete, de dobra de joelhos e de ondulações nas costas podem ser avaliados.

1.3 HIPÓTESE

Para resolver o problema, partiu-se da hipótese de que é possível detectar com sucesso o movimento completo cima/baixo do clarinete, o movimento de ondulação das costas e o movimento de dobra de joelhos presentes nos vídeos que compõem a base de dados de estudo, utilizando técnicas de visão computacional.

1.4 OBJETIVOS

1.4.1 Objetivo Geral

Este trabalho visa elaborar rotinas de visão computacional capazes de analisar vídeos de clarinetistas durante suas apresentações e, a partir somente destes, extrair informações de três movimentos auxiliares espontâneos característicos desta classe de músicos: movimento cima/baixo de clarinete, movimento de dobra de joelhos e movimento de ondulações nas costas, a fim de tornar o procedimento de análise de gestos mais eficiente.

Esta análise terá como base de dados nove vídeos obtidos apenas através da captura de som e imagem de quatro clarinetistas em um estúdio. Os músicos executam um trecho de cerca de 1 minuto do Segundo Movimento das Três Peças de Stravinsky para Clarinete Solo.

1.4.2 Objetivos Específicos

- Fazer a segmentação do músico, ou seja, eliminar a influência do cenário em que o mesmo se encontra.
- Determinar, nos vídeos, referenciais que sirvam de parâmetro para a detecção dos movimentos.
- Detectar movimentos auxiliares de transição cima/baixo do clarinete.
- Detectar movimentos auxiliares de dobra dos joelhos.
- Detectar movimentos auxiliares de ondulação das costas.
- Avaliar a eficiência dos algoritmos propostos frente a uma análise feita por um ser humano.

1.5 CONTRIBUIÇÕES

O método aqui descrito proporcionará resultados determinísticos relativos à detecção de movimentos específicos. A dependência humana intrínseca às visualizações será reduzida, devido ao método proposto resultar em indicadores matemáticos de existência ou não de tais movimentos auxiliares durante a execução da peça musical. Contribuir-se-á para o aprimoramento do atual processo de estudo comportamental dos clarinetistas durante a execução de uma música. Não será mais necessário dispor de muito tempo para detectar os movimentos apenas via observações, ou então trabalhar com muitos dados estatísticos na busca dos mesmos. Uma vez detectados, pode-se rotular os vídeos e trabalhar apenas nas regiões indicadas com presença de movimento.

Extrapolando o resultado aqui proposto, esta ferramenta será útil para pesquisadores que trabalham com o estudo comportamental humano sob um contexto musical. Entendendo como e por que o ser humano reage a determinados estímulos musicais, poder-se-ia gerar novas aplicações capazes de interpretar essas expressões, aprimorando todo o contexto. (AGGARWAL; CAI, 1999; GAVRILA, 1999; MITRA; ACHARYA, 2007).

1.6 ESTRUTURA

Esta dissertação é composta primeiramente por uma breve revisão bibliográfica, apresentando no Capítulo 2 o estado da arte relacionado à pesquisa. Em seguida, no Capítulo 3 é exposto o método de trabalho proposto, ou seja, os movimentos auxiliares em estudo são apresentados, juntamente com as técnicas de visão computacional utilizadas na análise. Dando sequência, o Capítulo 4 é composto pela avaliação dos resultados e o Capítulo 5 exhibe as conclusões obtidas. Finalmente no Capítulo 6 encontram-se todas as referências utilizadas durante a pesquisa.

2 ESTADO DA ARTE

Pesquisadores do mundo todo voltam seu foco para o estudo comportamental do ser humano no campo musical. Como e por que músicos agem durante suas apresentações são temas que estimulam pesquisas desde o campo da psicologia, para compreender melhor como as pessoas agem e reagem a manifestações musicais, como em áreas tecnológicas, na tentativa de extrair o máximo de informações possíveis oriundas desses movimentos, no intuito de aprimorar o espetáculo, ou criar novas atrações para o público (DELALANDE, 1988; CADOZ, 1988; WANDERLEY, 1999; DAHL; FRIBERG, 2004; GUNES; PICCARDI, 2007; JENSENIUS, et al., 2010).

De maneira geral, o aspecto visual torna as apresentações ainda mais atrativas. Os movimentos executados durante a música possuem expressões muito mais fortes que o próprio som (DAVIDSON, 1993). Este comportamento durante a execução musical leva diversas informações ao espectador, além da própria intenção sonora. Sentimentos como felicidade, tristeza, raiva e medo podem ser transmitidos à plateia através da movimentação dos músicos durante a execução (DAHL; FRIBERG, 2004).

Todo o estudo concentra-se na produção de movimentos, nos gestos, e estes possuem diversos conceitos. Em uma abordagem voltada à interação homem – máquina, gestos são movimentos corporais que contêm alguma informação (KURTENBACH; HULTEEN, 1990), ou que induzem a algum tipo de informação (HUMMELS; SMETS; OVERBEEKE, 1998).

Já sob o aspecto linguístico, gestos podem ser classificados como movimentos naturais das mãos gerados espontaneamente durante um discurso, ou uma conversa. São de caráter idiossincrático, ou seja, dependem de quem os executa, e questões comportamentais, educacionais e culturais influenciam diretamente na composição desses movimentos (WILSON; BOBICK; CASSELL, 1996).

Focado na abordagem de interesse – a musical –, inicialmente foi criado o conceito de “gestos instrumentais” (CADOZ, 1988). Neste estudo, Cadoz definiu diretrizes para o estudo específico de movimentos significativos à música

instrumental. Os gestos instrumentais são os movimentos aplicados a instrumentos, havendo interação física entre músico – instrumento, e subdividem-se em gestos de excitação, de modificação e de seleção (CADOZ, 1988).

Este conceito aplica-se única e exclusivamente à produção sonora. No entanto, durante a execução da música, outros movimentos podem ser observados, mesmo não produzindo sons (WANDERLEY, 1999). Nessa linha de raciocínio, Delalande, em suas observações com um pianista, instituiu a classe de “gestos musicais”. Esta coleção de movimentos foi subdividida em três classes: gestos efetivos, gestos de acompanhamento e gestos figurativos (DELALANDE, 1988).

Na concepção de Delalande, os gestos efetivos, como o próprio nome sugere, são os movimentos que efetivamente produzem som, como, no caso, os dedos pressionando as teclas do piano. Os gestos de acompanhamento, também de modo sugestivo, são os movimentos que acompanham o processo de produção musical, como movimentos dos ombros e da cabeça. Por fim, os gestos figurativos são todos aqueles que são perceptíveis pela mudança sonora, mas não apresentam correspondência com movimentos em si – normalmente gerados por mudanças na melodia ou da entonação – (DELALANDE, 1988).

Ambas as classificações foram adaptadas à realidade dos clarinetistas. Em seus estudos, Wanderley primeiramente propôs duas classificações de gestos para esses instrumentistas: gestos livres, onde não há contato físico entre músico e instrumento, e gestos interativos, onde há algum tipo de contato entre ambos (WANDERLEY, 1999).

Aplicando a teoria de gestos instrumentais de Cadoz ou de gestos efetivos de Delalande nesta configuração de Wanderley, pode-se descrever o comportamento de um clarinetista durante a execução de uma música. Entretanto, o estudo fica apenas centrado na produção sonora.

Como já fora citado, outras ações além da produção musical são observadas e não se encaixam nessas configurações. Esses gestos são importantes para o aspecto visual da música, uma vez que os movimentos executados possuem expressões muito mais fortes do que o próprio som (DAVIDSON, 1993).

Baseado nisso, aplicou-se, então, o conceito de gestos de acompanhamento de Delalande no universo dos clarinetistas, mas primeiramente focando apenas na região de atuação do instrumento. Estudaram-se, então, movimentos como

elevação, abaixamento, de um lado ao outro e vibração do clarinete. Estes gestos passaram a ser tachados de “gestos auxiliares” ou “gestos não óbvios”. (WANDERLEY, 1999).

Esta nomenclatura está intimamente ligada à natureza dos movimentos. Justamente por não produzirem sons, ou seja, por aparentemente não terem significado algum em sua execução, acabam sendo caracterizados como não óbvios. Ademais, é perfeitamente possível executar uma música de forma estática, porém, esta normalmente caracteriza-se por um estado forçado, não espontâneo. De maneira natural, os movimentos de acompanhamento aparecem durante as execuções musicais. Logo, auxiliam no processo de criação musical.

Embora não tenham o claro objetivo de produzir som, eventualmente implicam em pequenas modificações sonoras (WANDERLEY, 1999; WANDERLEY, 2002). Experimentos com diversos microfones captando algumas apresentações de clarinetistas comprovaram que gestos não óbvios de fato não produzem sons, mas eventualmente causam alterações de amplitude sonora no resultado final, pois a presença de movimentos causava mudanças no direcionamento do som, influenciando em parâmetros como reverberação ou eco no ambiente (WANDERLEY, 1999).

Na tentativa de sistematizar o estudo, os movimentos auxiliares foram também subdivididos conforme sua origem. Surgiram três novas categorias: movimentos fisiológicos, estruturais e interpretativos. Os fisiológicos estão intimamente ligados à pessoa, como, por exemplo, a respiração ou movimentos das mãos e dedos. Os estruturais têm uma relação direta com a composição musical, com a estrutura da música e os interpretativos dependem da entonação que o instrumentista pretende dar à música, do sentimento a ser passado. Logo, são os mais idiossincráticos e difíceis de serem observados (WANDERLEY, 2002).

Baseado nesta nova abordagem, análises indicam outros e curiosos resultados. Aplicando a teoria de movimentos estruturais, observa-se que há certa tendência dos músicos a mudarem de postura no início das frases (trechos musicais). Já, no decorrer das frases, a tendência é de apresentarem movimentos longos e contínuos durante notas longas e movimentos curtos e frenéticos durante muitas notas curtas, ou na presença de muitas notas (WANDERLEY, 2002; WANDERLEY, et al., 2005).

Além disso, os mesmos músicos tendem a um padrão gestual quando executam o mesmo excerto em momentos distintos, conforme as partituras. Há movimentos indiferentes aos músicos enquanto outros aparentam ser intimamente dependentes deles (CADOZ; WANDERLEY, 2000; WANDERLEY, 2002).

Conforme já citado, alguns movimentos auxiliares são tachados como inerentes à classe. São eles o movimento completo cima/baixo do clarinete, movimento circular completo do clarinete, movimento cima/baixo de cabeça, movimento cima/baixo de ombros, movimento de ondulação das costas, movimento de abertura/fechamento de braços (“bater asas”), movimento de inclinação da cintura, movimento de dobra de joelhos, movimento de inclinação e batida de pés, deslocamento de peso para a esquerda ou para a direita. (WANDERLEY, et al., 2005).

Dentre esta listagem, o gesto pendular (cima/baixo) do clarinete, ondulação das costas, dobra dos joelhos e deslocamento de peso são os mais frequentes. Além disso, são movimentos com alto grau de distinção entre si e por isso são bastante utilizados em estudos desta natureza (WANDERLEY, et al., 2005; VERFAILLE, et al., 2006 ; JENSENIUS, et al., 2010). Por exemplo: a sonificação de movimentos – uso de áudio para transformação de dados de gestos em percepções de gestos, a fim de facilitar a comunicação e a interpretação – (VERFAILLE, et al., 2006), no intuito de criar outra perspectiva de compreensão dos movimentos auxiliares, utiliza estes quatro movimentos como base de estudo.

Além da sonificação, outras técnicas de análise são empregadas no intuito de compreender a forma do ser humano de agir perante gestos. Em uma abordagem de interface homem-máquina, uma proposta para estudo de gestos compreende o rastreamento de pontos-chave obtidos via funções de curvatura associadas às trajetórias (BANDERA, J.; MARFIL; BANDERA, A.; RODRÍGUEZ; MOLINA-TANCO; SANDOVAL, 2009).

Outro método para avaliação de movimentos é o de Modelo Temporal, composto por um vetor-imagem onde cada componente deste é uma função de propriedades de movimentos correspondentes à localização espacial numa sequência de imagens (BOBICK; DAVIS, 2001).

Redes neurais também têm um papel de destaque nas pesquisas, como o uso destas numa metodologia de reconhecimento de expressões faciais

(ROSENBLUM; YACOOB; DAVIS, 1996).

Voltando ao foco dos clarinetistas, o método mais comum é o uso de marcadores. Através desse procedimento, pode-se aplicar um método reverso de movimento, no intuito de tornar virtual o músico em estudo (MAZZARINO, et al., 2009), ou utilizar os dados de captura para a criação de correlações entre quantidades de movimento e a estrutura musical que está sendo executada (WANDERLEY, et al., 2005; TEIXEIRA, 2010), ou ainda simplesmente analisar as trajetórias durante a execução da peça (WANDERLEY, 1999; WANDERLEY, 2002).

Todas essas metodologias são propostas para a detecção ou mensuração de movimentos de forma computacional, a fim de criar ferramentas para o estudo comportamental do ser humano. Em suma, muito já fora estudado, entretanto ainda há um grande campo a se explorar (CADOZ; WANDERLEY, 2000). Aqui se propõe um método alternativo para a detecção de movimentos auxiliares de clarinetistas com este mesmo propósito: uma ferramenta cujos resultados auxiliem no estudo comportamental humano.

3 MÉTODO PROPOSTO

3.1 VISÃO GERAL

Para poder analisar a base de vídeos e a partir destes mensurar os três movimentos auxiliares neles contidos, foi criado um método composto por três partes fundamentais: pré-processamento dos vídeos, detecção de movimentos e avaliação dos resultados, conforme ilustra o diagrama de blocos da Figura 5.

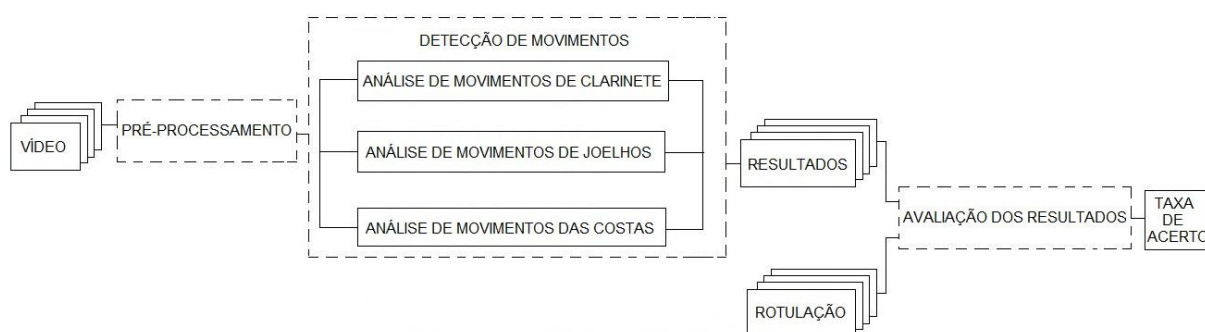


Figura 5: Visão geral do método proposto.

Qualquer avaliação computacional de movimentos partindo da base de dados original é muito complexa. Há muita informação desnecessária ao estudo, gerando erros ou até mesmo inviabilizando o uso de determinadas técnicas de processamento de imagem. Além disso, o excesso de informação exige maior recurso computacional para gerar bons resultados. Por esse motivo que o pré-processamento é muito importante.

Esta etapa inicial simplesmente adéqua o ambiente de trabalho. Tudo que é “estranho” ao processo é devidamente filtrado e o restante da informação é ajustado com o intuito de potencializá-la, aprimorando os dados e facilitando o estudo para as etapas seguintes.

Como o estudo será focado nos movimentos, um bom pré-processamento indica bons dados para avaliações posteriores. A eliminação ou meramente a atenuação de diversos agentes indutores de erro é de suma importância para o processo, aumentando o grau de precisão e certeza das tomadas de decisão que se

fazem necessárias.

As tomadas de decisão estão presentes na detecção de movimentos. Nesta fase, os três movimentos auxiliares sugeridos serão avaliados. Para isto, cada vídeo pré-processado com o músico executando uma peça passará por um arranjo de três blocos, cada um com seus algoritmos particulares capazes de detectar a presença ou não do respectivo movimento.

A etapa de detecção de movimentos pode ser considerada um pós-processamento. Tudo que há nela depende fortemente do que fora obtido pelo pré-processamento e resultará em parâmetros e indicativos fundamentais à proposta de avaliação dos movimentos.

Finalmente, a avaliação de resultados encerra o processo. Todos os resultados obtidos através de decisões da etapa imediatamente anterior são então avaliados, a fim de se mensurar a taxa de acerto dos mesmos.

Taxa de acerto pode parecer um termo inadequado, pois movimentos são determinísticos – existem ou não. Todavia, os algoritmos de detecção trabalham de forma relativa, mais de um parâmetro é avaliado no intuito de caracterizar determinado gesto. O grau de influência de cada parâmetro implica numa robustez diferente do sistema, influenciando na taxa final de acerto.

Comparando os resultados fornecidos pelo algoritmo com os esperados, via outros métodos de análise, é que se pode realmente definir o índice de confiabilidade da pesquisa e a etapa de avaliação de resultados que é responsável por isso.

A seguir serão detalhadas as etapas de pré-processamento e detecção de movimentos. O item avaliação de resultados será discutido no Capítulo 4.

3.2 BASE DE DADOS

Primeiramente resume-se na Tabela 1 a base de dados trabalhada na pesquisa.

Tabela 1: Base de vídeos.

VÍDEO	FORMATO	DURAÇÃO	TAMANHO	RESOLUÇÃO	TAXA DE QUADROS	CODEC
1.1	MOV	1min13s	39.6MB	300 x 400	15 fps	CVID
1.2	MOV	1min11s	38.4MB	300 x 400	15 fps	CVID
1.3	MOV	1min27s	46.9MB	300 x 400	15 fps	CVID
1.4	AVI	1min09s	7.46MB	480 x 720	25 fps	XVID
2.1	MOV	1min18s	42.5MB	300 x 400	15 fps	CVID
2.2	MOV	1min31s	49.5MB	300 x 400	15 fps	CVID
2.3	MOV	1min19s	43MB	300 x 400	15 fps	CVID
3	AVI	1min22s	8.88MB	480 x 720	25 fps	XVID
4	AVI	1min07s	7.28MB	480 x 720	25 fps	XVID

Conforme mencionado no Capítulo 2 e ilustrado na Tabela 1, a base de dados é formada por nove vídeos gerados por quatro clarinetistas. Quatro vídeos são de um clarinetista, ao passo que três outros vídeos são de um segundo músico. Outros dois instrumentistas completam a base.

3.3 PRÉ-PROCESSAMENTO

O principal foco do pré-processamento foi eliminar a influência do cenário no vídeo e corrigir problemas de iluminação durante as filmagens. É de fundamental importância para as futuras detecções, uma vez que o mesmo é responsável direto pela adequação da base de dados à capacidade de reconhecimento dos algoritmos propostos para os gestos em estudo. Dificuldades nesta etapa inicial podem acarretar dificuldades na detecção de movimentos.

O diagrama de blocos exibido na Figura 6 ilustra o procedimento adotado para o pré-processamento da base de vídeos.

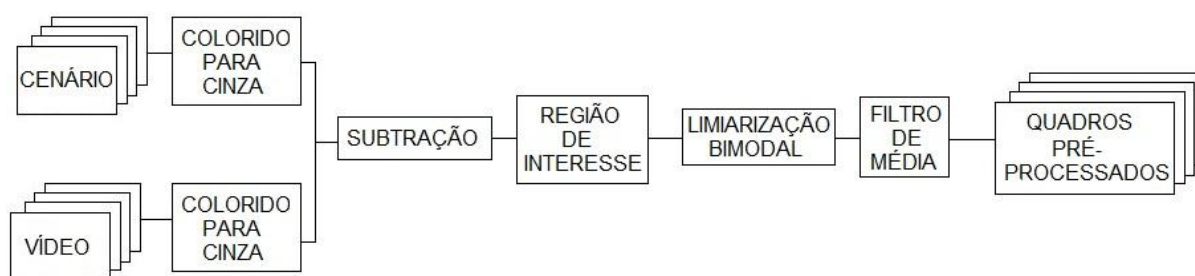


Figura 6: Diagrama de blocos da etapa de pré-processamento.

Os nove vídeos da base de dados eram compostos de um único cenário ilustrado na Figura 7.



Figura 7: Vídeo original.

A técnica de visão computacional escolhida para a eliminação do cenário foi a de subtração de fundo (*background subtraction* - BS) (SZELISKI, 2010). Este procedimento foi escolhido devido ao ambiente ser um estúdio, ou seja, ser um ambiente controlado e estático. Controlado, pois a iluminação é quase que constante, não há mudança brusca de quantidade de luz e estático, porque o cenário é fixo, apenas o músico se desloca, enquanto que a câmera e o ambiente em si permanecem imóveis. Dessa forma, o BS possui aplicação viável.

Para a utilização da técnica, faz-se necessário o armazenamento de um quadro de referência que será subtraído dos quadros em análise. Como o intuito é eliminar o cenário, ou seja, tudo além do músico, o quadro do vídeo contendo apenas o ambiente, sem o músico, é a referência ideal.

A subtração de fundo se torna trabalhosa pelo fato de os vídeos serem coloridos. As três camadas de cor do sistema RGB complicam o processamento, pois a variação dos pixels de quadro para quadro é considerável, gerando muito ruído no resultado pós-BS. A solução encontrada foi aplicar uma mudança inicial no mapeamento de cores, transformando os vídeos para níveis de cinza com 8 bits de profundidade. Além disso, no intuito de destacar o músico, foi aplicada uma inversão de pixels, tornando o cenário mais escuro que o instrumentista.

Estas transformações de cores, além de facilitarem o processamento dos vídeos, ajudam também a reduzir o tamanho destes. Outro fator que implica na redução dos arquivos é a possibilidade de filtragem do canal de áudio, pois este não é necessário à análise. Uma vez preparados e com o quadro de cenário identificado, pôde-se, enfim, aplicar o BS nos vídeos.

Ao passo que o BS faz uma subtração de quadros do vídeo por um quadro de cenário fixo, a região que compreende este cenário tende a tons de preto (pixel 0) ou branco (pixel 255), enquanto que a localização do músico gira em tons medianos de cinza, conforme ilustra a Figura 8.



Figura 8: Subtração de fundo.

Como a análise é em nível de pixels, esta gera imprecisões, ou ruídos. Se a imagem fosse 100% homogênea, como à primeira vista parece ser, toda a região imóvel ficaria escura. No entanto, mínimas variações de pixels que não são vistas a olho nu são maximizadas ao aplicar-se tal matemática. Quando há resultado positivo da subtração de pixel, ou seja, o quadro em análise possui um tom de cinza igual ou levemente superior ao da referência, o resultado tende a 0, porém, no caso inverso, quando o tom de cinza é levemente inferior à referência e, por conseguinte, a subtração gera um falso pixel negativo, a mesma operação matemática leva este pixel a níveis de branco. Isso explica a ocorrência deste aspecto na Figura 8. Fontes

de iluminação, sombras e má qualidade de captação de imagem podem gerar estas imperfeições.

Para eliminar este efeito indesejável, definiram-se, primeiramente, regiões de interesse. Isto poderia ter sido feito já inicialmente, colaborando para a redução da quantidade de informação a ser pré-processada, no entanto, tal necessidade só foi percebida logo que a subtração de fundo gerou os resultados mostrados na Figura 8.

Todos os nove vídeos da base de dados foram editados a fim de cortar o máximo possível do cenário, sem perder informações dos instrumentistas. Pôde-se executar esta operação pelo fato de os músicos não promoverem grandes deslocamentos pelo cenário, uma vez que necessitavam ler a partitura para executar o excerto.

Excluindo tudo que não pertencesse à área de atuação do clarinetista, obteve-se o resultado mostrado na Figura 9.



Figura 9: Região de interesse.

Outro procedimento aplicado no pré-processamento dos vídeos foi o da limiarização bimodal da imagem (SHAH, 2010; SZELISKI, 2010). Uma vez que o músico e seu clarinete se movimentam, acabam sendo formados por tons medianos de cinza, enquanto que cenário e ruído são tons de cinza próximos aos limiares 0 e

255. Aplicou-se, então, uma função de transferência à imagem, através de uma simples lógica condicional: os pixels da imagem que fossem próximos de 0 ou de 255 foram transformados em nível 255, já a faixa intermediária restante foi transformada em 0. Ademais, foi aplicada uma suavização da imagem por filtro de média, com máscara 3x3 (GONZALEZ; WOODS, 2007), a fim de aprimorar a definição dos contornos.

Aplicada esta metodologia, conseguiu-se, enfim, segmentar o clarinetista de todo o cenário, além de eliminar a influência da iluminação. A Figura 10 ilustra o pré-processamento aplicado aos vídeos de cada um dos quatro instrumentistas.

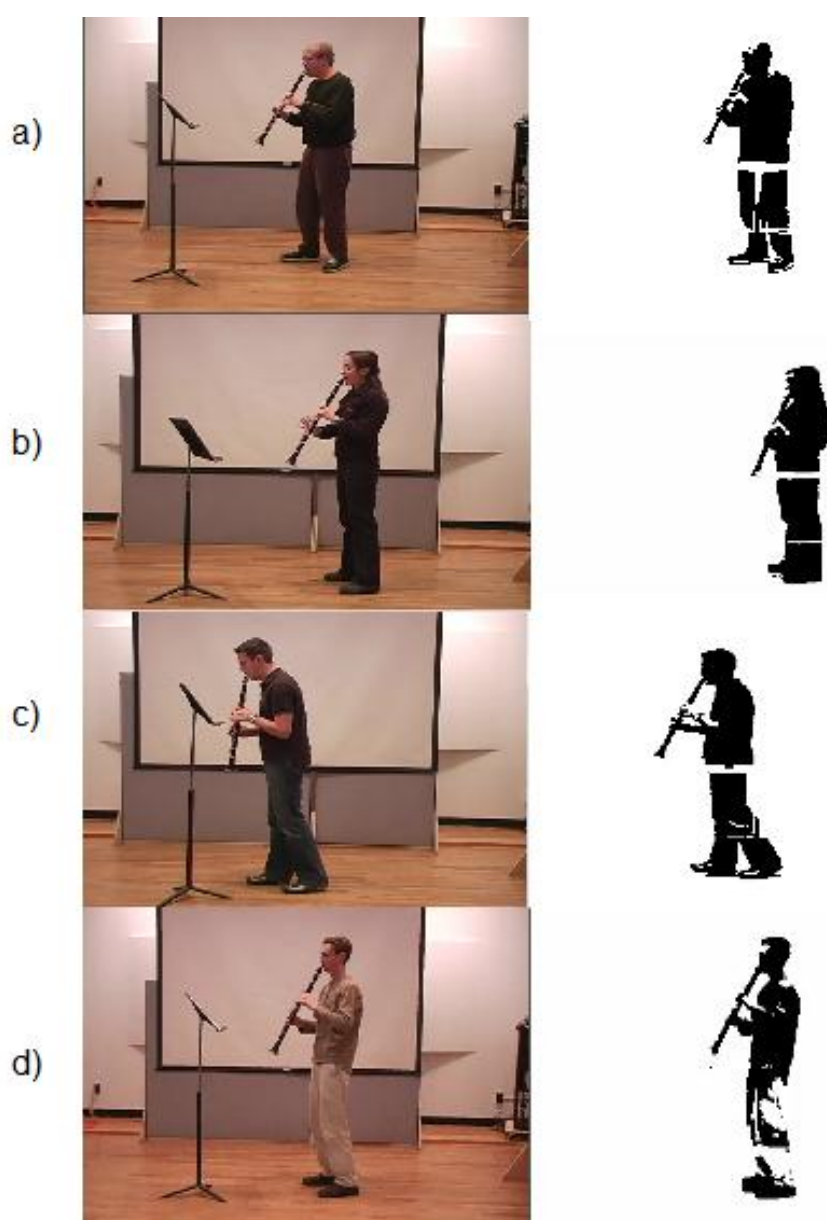


Figura 10: Comparação entre vídeo original e vídeo pré-processado.

Os vídeos do clarinetista 1 (Figura 10a), do clarinetista 3 (Figura 10b) e do clarinetista 4 (Figura 10c) tiveram pré-processamento melhor que os vídeos do clarinetista 2 (Figura 10d). Pode-se dizer que o cenário foi eliminado, os problemas de iluminação foram atenuados e o músico foi segmentado por completo. O clarinetista 2 não foi perfeitamente segmentado pelo fato de usar roupas claras. Isso dificultou o uso da técnica de limiarização bimodal da imagem, gerando essas imperfeições.

Vale a pena ressaltar que não há como obter novos vídeos para corrigir eventuais falhas de posicionamento e iluminação, ou ainda refazer as filmagens especificando a roupa adequada que o músico deveria utilizar. A base de dados é fixa e única. Portanto, o pré-processamento proposto é viável e gera bons resultados somente quando há um quadro de referência – cenário – e que os instrumentistas apresentem forte contraste com ele.

Concluída esta etapa, o foco passa então para a detecção de movimentos.

3.4 ROTULAÇÃO

Antes do reconhecimento de movimentos propriamente dito, foi necessário criar um método para avaliar o desempenho dos algoritmos de reconhecimento propostos, ou seja, predefinir quais movimentos ocorrem e em quais quadros do vídeo eles estão presentes – rotular os vídeos. Desse modo, basta comparar os quadros indicados pelo método proposto com esta rotulação e computar a taxa de acerto.

O principal objetivo da pesquisa é substituir a visualização dos vídeos pela detecção auxiliada via algoritmos computacionais. Entretanto, para avaliar os resultados destes algoritmos, faz-se necessária a comparação dos mesmos com o método clássico, ou seja, com as detecções de movimentos a partir de observações dos vídeos. Somente desta forma que se poderia afirmar que o método proposto é plausível e seria capaz de substituir o método clássico e, com isso, colaborar com o estudo de movimentos auxiliares. Baseado nisso que foi decidido criar um simples *software* para colaborar com as observações.

Este *software* é responsável simplesmente por transformar a análise visual em uma tabela formada pela contagem de todos os quadros que compõem o vídeo e

os gestos alvos de estudo. A Figura 11 ilustra a interface deste *software*.



Figura 11: Rotulação dos vídeos.

A interface é simples: uma janela de exibição do vídeo, três botões para cada um dos movimentos e um botão “Análise”, para gerar o resultado da rotulação.

Ao observar um movimento, o usuário utiliza o botão adequado para indicar quando determinado movimento começou e quando terminou, e o programa faz a anotação dos quadros deste intervalo. Ao final do vídeo, a rotina de análise gera um arquivo de rotulação para este vídeo, contendo a disposição quadro a quadro de todos os movimentos observados.

Como fora comentado, o algoritmo apenas auxilia na geração de dados que serão referência para a análise, não havendo nenhuma assistência à detecção de gestos. Assim, não impede a ocorrência de problemas na rotulação.

O tempo de reação entre visualização do movimento e o manuseio do *mouse* para indicação do mesmo, por exemplo, culmina com distorções temporais na detecção dos gestos, ou atrasos de detecção. Outro exemplo é a possível não observação de movimentos sutis e/ou de curta duração, determinando a ausência de gestos na tabela criada pelo *software*. Esses problemas serão comentados no Capítulo 4.

3.5 DETECÇÃO DE MOVIMENTOS DE INTERESSE

É a etapa responsável pela detecção de cada um dos gestos auxiliares em questão. Cada bloco desta repartição trabalhará com elementos do vídeo pré-processado no intuito de identificar movimentos completos cima/baixo do clarinete, movimentos de dobra dos joelhos, e, por fim, movimentos de ondulações nas costas.

3.5.1 Movimentos de Clarinete

Dentre os três movimentos propostos para análise, os gestos de clarinete aparecem como mais evidentes nos vídeos. Como se trata de um instrumento anexo ao corpo do clarinetista, sua identificação é quase que imediata.

A proposta para detecção dos movimentos de clarinete durante a execução dos vídeos foi basicamente de rastrear o instrumento. Isto seria análogo ao procedimento de aquisição de dados via marcadores, contudo, puramente via técnicas de visão computacional usando marcadores virtuais, sem o uso de tais dispositivos. O diagrama exibido na Figura 12 ilustra o procedimento.

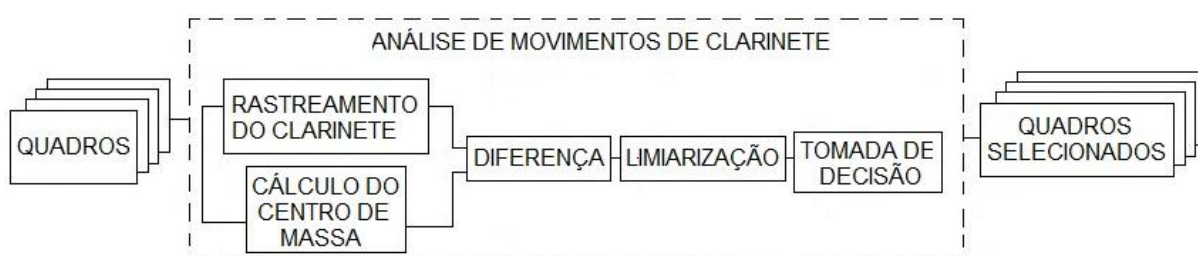


Figura 12: Diagrama de blocos do método de análise de movimentos de clarinete.

O movimento cima/baixo do clarinete, também conhecido como movimento pendular, pode ser completamente descrito em função de duas coordenadas: uma vertical e outra horizontal. O método proposto foi, então, a alocação de um marcador virtual na extremidade do clarinete, logo nos primeiros quadros do vídeo em execução. Ao passo que os quadros progridem e o clarinetista produz movimentos na região, o marcador virtual acompanha a trajetória do instrumento, armazenando as suas coordenadas em um vetor para posterior análise.

O algoritmo utilizado para a o rastreamento da trajetória do marcador virtual foi baseado no programa LkDemo, um simples algoritmo com este propósito

desenvolvido em linguagem C, com o auxílio da biblioteca específica para visão computacional: o OpenCV (BRADSKI; KAEHLER, 2008). Através deste software pode-se inserir um marcador virtual e o algoritmo faz o rastreamento deste ponto a partir da estimação de estado durante a transição de quadros através do método diferencial de Lucas-Kanade para estimação de fluxo óptico (LUCAS; KANADE, 1981). O método de Lucas-Kanade assume que o deslocamento de um *pixel* entre dois quadros da imagem é pequeno e aproximadamente constante dentro de sua vizinhança. Baseado nisso, o algoritmo fornece uma estimativa da coordenadas x, y deste mesmo *pixel* para todos os quadros subsequentes do vídeo.

O objetivo deste procedimento era meramente possibilitar ao operador a inserção de um marcador virtual e ter então as coordenadas deste ponto armazenadas em um vetor, quadro a quadro. Como o movimento pendular é um movimento brusco, visível, gera uma grande variação do valor da coordenada do marcador, ou seja, o rastreamento não precisa ser rigorosamente exato, pequenos erros de precisão são toleráveis. Dessa forma, o algoritmo proposto, apesar de simples, foi o suficiente. A Figura 13 ilustra o rastreamento do marcador virtual.

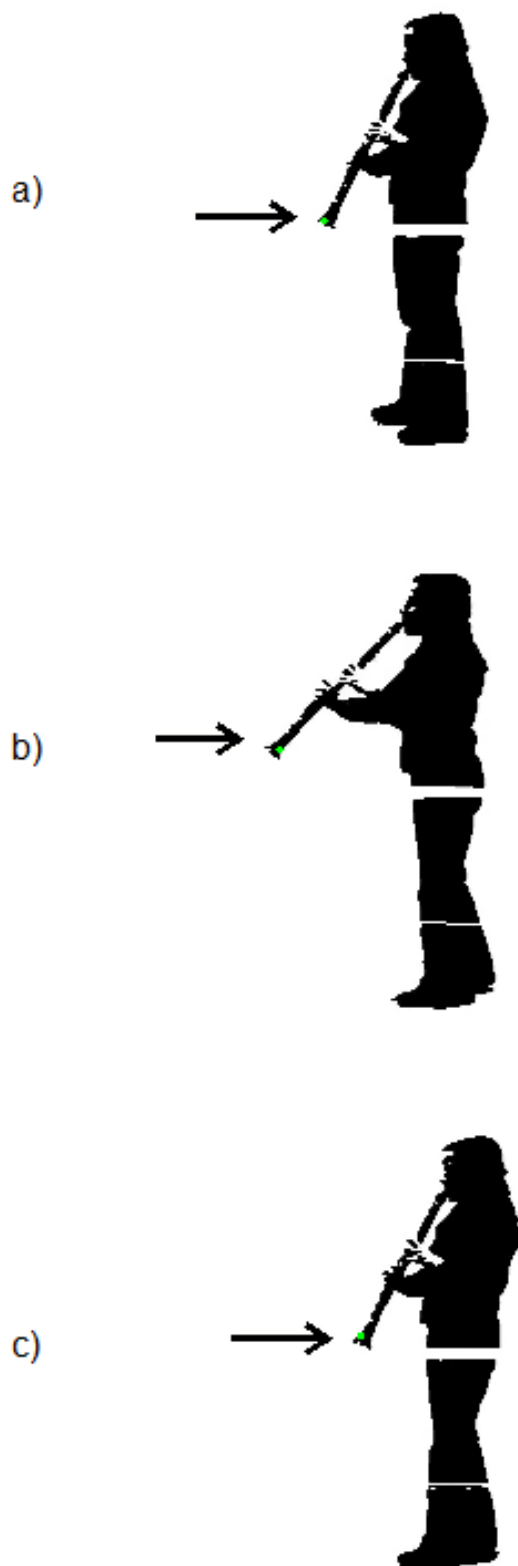


Figura 13: Exemplo de deslocamento do marcador virtual rastreado.

Observando a Figura 13, nota-se que há um grande deslocamento do ponto na sequência de quadros a, b e c. Isso pode ser traduzido para um gráfico do rastreamento, conforme visto na Figura 14.

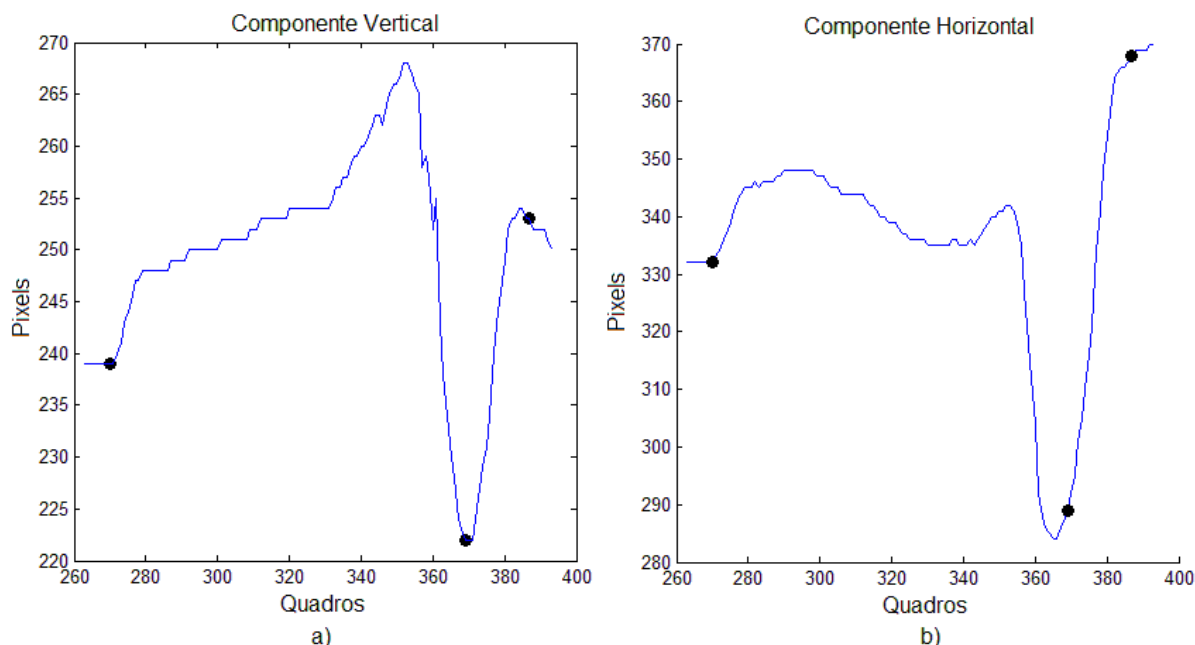


Figura 14: Trajetória do marcador virtual rastreado: a) componente vertical; b) componente horizontal.

Vale a pena ressaltar que em análises de *pixels*, o ponto (0,0) localiza-se na extremidade superior esquerda do vídeo, e não na inferior esquerda, como numa representação cartesiana. Portanto, como o movimento do clarinete é evidentemente para cima, isso significa que a coordenada do marcador virtual em seu ápice ficou o mais próxima da origem, caracterizando os vales nos gráficos da Figura 14.

A Figura 14a mostra a variação da componente vertical, enquanto que a Figura 14b mostra o deslocamento da componente horizontal. Ambas sofrem grande variação, ou seja, as duas possibilitam a caracterização do movimento. Baseado nisso, foi possível simplificar o método de análise para a observação de apenas uma delas. Ao observar-se o movimento pendular do clarinete, este se destaca por um deslocamento na vertical. Por isso, optou-se por analisar apenas a componente vertical das coordenadas obtidas pelo método de rastreamento.

Como fora comentado em capítulos anteriores, medir um movimento é algo relativo. Simplesmente analisar a componente vertical do marcador virtual não é suficiente para se tomar decisões acerca da presença ou não dos gestos porque outros fatores podem influenciar na mudança destas coordenadas. É preciso ter um referencial comparativo, algo capaz indicar se tal trajetória rastreada representa ou não um movimento. É preciso relativizar esta medida a partir de um referencial que se desloque da mesma forma, junto com o músico. Por isso, decidiu-se calcular o centro de massa dos instrumentistas.

O centro de massa é adequado porque se trata de um referencial interno ao instrumentista, e não ao cenário geral. Utilizar um referencial externo implicaria em problemas quanto ao deslocamento do músico pelo ambiente. O centro de massa desloca-se proporcionalmente ao músico. Ao abaixar-se, o centro de massa também se move para baixo, assim como quando o músico caminha, o ponto também acompanha, garantindo a estabilidade e coerência do referencial.

Em outras pesquisas relacionadas a reconhecimento de gestos, Li e Greenspan sugeriram o uso desta técnica para criar assinaturas de movimentos. Nela, mediam a distância de um ponto centroide a pontos nas mãos e geravam curvas de magnitude de movimentos. Estas futuramente serviam para criar as assinaturas (LI; GREENSPAN, 2011). O movimento de clarinete se adéqua a este contexto.

Uma vez que os vídeos não possuem mais cenário, ou seja, só há músico e clarinete, calcular o centro de massa do instrumentista significa calcular o centro de massa de todo o quadro. Pode-se simplificar este cálculo justamente pelo fato de ter-se garantido que onde há *pixels* pretos, há a pessoa. Partindo deste pressuposto, uma média dos *pixels* no sentido horizontal e vertical resulta na coordenada do centro de massa para cada quadro, segundo a equação 1.

$$cm^q = \left(\frac{\sum_1^w \sum_1^h x}{N}, \frac{\sum_1^w \sum_1^h y}{N} \right) \quad (1)$$

Onde “q” é o quadro em análise, “x” é a componente vertical da coordenada do *pixel* preto, “y” é a componente horizontal da coordenada do *pixel* preto, “h” é a componente vertical do tamanho do quadro “q”, “w” é a componente horizontal do tamanho do quadro “q” e N é quantidade de *pixels* pretos no quadro “q”.

A Figura 15 ilustra a posição do centro de massa para alguns quadros do vídeo.

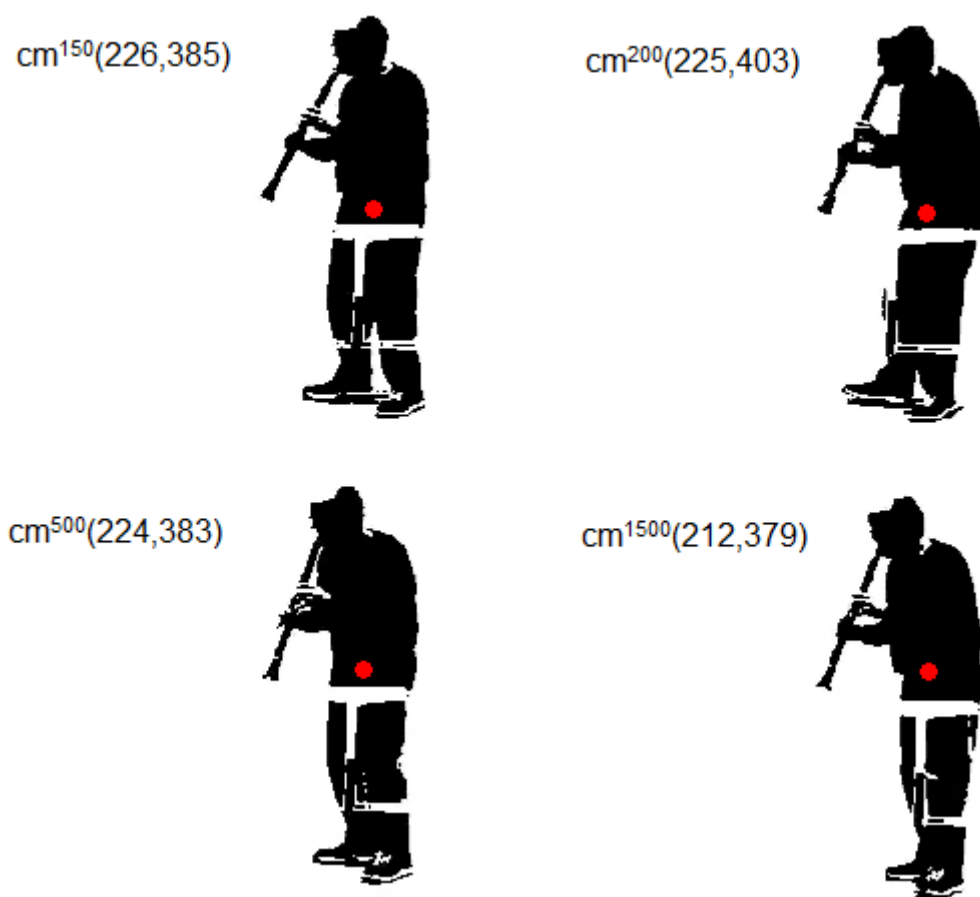


Figura 15: Posição do centro de massa ao longo de quadros do vídeo.

Novamente surge a questão da precisão do cálculo. Como há imperfeições na imagem, como é possível observar na Figura 15, a média dos *pixels* não retorna o valor exato do centro de massa. Entretanto, assim como no caso do marcador virtual, não se faz necessário o uso do ponto com extrema precisão. A relação entre as coordenadas acaba sendo maior do que estes pequenos erros.

Tendo armazenados os dados de deslocamento do marcador e do centro de massa, pode-se partir para a etapa de pós-processamento, ou seja, trabalhar com os dados a fim de proporcionar a detecção de movimentos.

A caracterização do movimento passou a ser uma relação da trajetória do ponto conhecido no clarinete com a trajetória do centro de massa. A diferença absoluta entre essas duas coordenadas sugere, agora, a presença ou não do gesto. Um valor alto de diferença indica um afastamento, ao passo que um valor baixo representa proximidade dos pontos. É essa transição entre alto – baixo que descreve o gesto auxiliar de clarinete.

A diferenciação entre a trajetória do ponto-referência e o centro de massa gera uma nova curva relativa aos dois deslocamentos. Picos nesta nova curva sugerem grandes distâncias entre as duas primeiras, ou seja, afastamento, enquanto que vales indicam proximidade entre as mesmas, ou a proximidade dos pontos. Esta curva poderia ser gerada automaticamente ao se referenciar o rastreamento da trajetória do marcador virtual diretamente ao centro de massa, no entanto, como o centro de massa também será útil aos outros métodos de detecção de movimentos, optou-se por calculá-lo separadamente, gerando um arquivo contendo as coordenadas em cada quadro para cada um dos nove vídeos. A Figura 16 exemplifica uma curva de diferença sobreposta a sua rotulação.

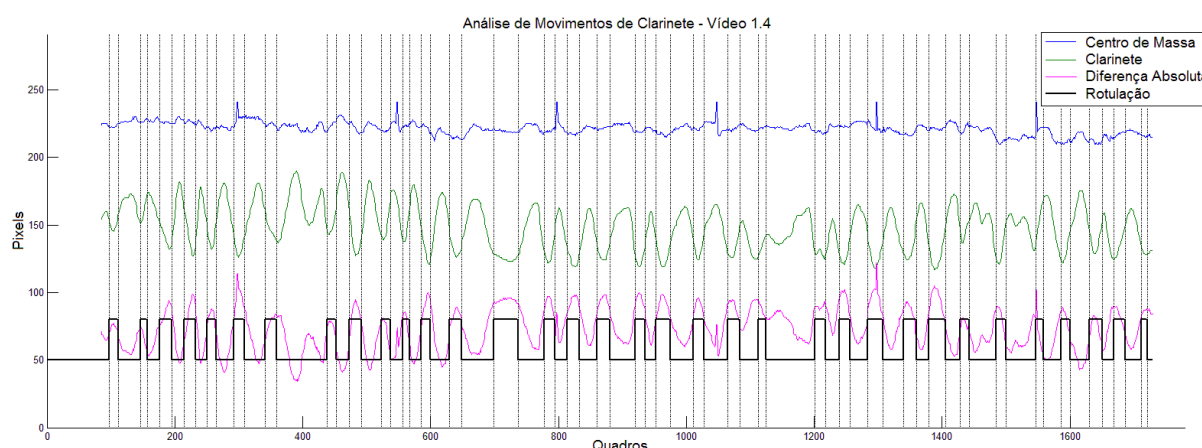


Figura 16: Exemplo de curva de diferença sobreposta à rotulação.

Gerada a curva de diferença, partiu-se para a limiarização da mesma. Diferentemente do processo de limiarização bimodal aplicado no pré-processamento dos vídeos, este procedimento agora é uma classificação, uma espécie de comparador. Baseado em um limiar L , particular de cada vídeo analisado, sempre que o valor da curva de diferença for igual ou superior a este limite, eleva-se a curva para um nível 1, ao passo que quando o valor for inferior a L , rebaixa-se a 0. Isso foi feito, pois, reparando na Figura 16, nota-se que há similaridade entre instantes de movimento sugeridos pela rotulação e formação de máximos locais na curva. Explica-se esse padrão pelo fato de o movimento de clarinete ser visivelmente notado quando o instrumento está na proximidade do ápice de movimento.

Aplicando a limiarização proposta, cria-se um aspecto digital da curva de diferença, similar à curva de rotulação, transformando a informação de distância

entre os pontos de referência em uma representação de presença ou ausência de movimento durante a sucessão de quadros dos vídeos.

Como os vídeos não são idênticos – apresentam características particulares – o fator L aplicado às curvas de diferença também não é padrão. A aproximação matemática relacionada a ele pode gerar falsos positivos (detecção de movimentos quando, na verdade, não há) quando é pequeno, mas quando muito grande, pode acarretar em falsos negativos (não detecção de movimentos), principalmente os gestos de menor amplitude.

Para tornar o fator limitador um número intrínseco às propriedades dos vídeos, optou-se por gerá-lo a partir do estudo estatístico de cada curva de diferença gerada. Dessa forma, obteve-se uma combinação do valor médio (d_M) e o desvio padrão (d_{DP}), gerando os limiares exibidos na Tabela 2.

Tabela 2: Limiares da curva de diferença absoluta para cada vídeo.

VÍDEO	d_M	d_{DP}	L
1.1	24	13	$d_M + 0.5*d_{DP} = 31$
1.2	22	6	$d_M + 0.5*d_{DP} = 25$
1.3	18	11	$d_M + 0*d_{DP} = 18$
1.4	73	15	$d_M + 0*d_{DP} = 73$
2.1	3	2	$d_M + d_{DP} = 5$
2.2	9	6	$d_M + d_{DP} = 15$
2.3	21	9	$d_M + d_{DP} = 30$
3	12	9	$d_M + d_{DP} = 21$
4	28	11	$d_M + d_{DP} = 39$

Ao se observar os valores de L da Tabela 2, nota-se que diferentes valores foram utilizados para o mesmo clarinetista. A explicação para este fato é o de cada vídeo apresentar uma forma diferente de execução do excerto. Há vídeos em que o músico age normalmente, enquanto que outros foram obtidos em modo forçado, seja para execução com grande ou nenhuma expressividade. Isso afeta diretamente no reconhecimento de gestos. Movimentos exagerados toleram limiares maiores, enquanto que gestos mais singelos exigem limiares menores.

Os limiares apresentados na Tabela 2 foram utilizados no intuito de restringir ao máximo o número de falsos negativos. Falsos positivos também não são ideais, mas, como o objetivo do trabalho é substituir a visualização dos vídeos, uma simplificação do atual método, para que os movimentos sugeridos sejam

posteriormente avaliados via outros recursos, esses erros são toleráveis. Já as não detecções implicariam em perda de dados para os estudos subsequentes.

Aplicados os conceitos aqui exibidos, foram obtidos dados relativos à detecção de movimentos de clarinete para cada um dos nove vídeos. Estes resultados serão exibidos e comentados na Seção 4.

3.5.2 Movimentos de Dobra de Joelhos

Diferentemente da análise de clarinete, o movimento de dobra de joelhos não permite o rastreamento de ponto para sua verificação. Isso ocorre, em suma, por dois motivos: primeiro porque a região do joelho não é uma área destacada do corpo. O clarinete é um instrumento anexo ao músico, quase que isolado dele, o que permite o rastreamento. Entretanto, os joelhos são parte integrante do corpo do instrumentista, corpo este que após o pré-processamento torna-se um único elemento, sem subdivisões claras. Dessa forma, o algoritmo de rastreamento perde o referencial facilmente entre as transições de quadros.

O segundo motivo para não se aplicar o rastreamento de ponto nesta ocasião é pelo fato do gesto de dobra de joelhos ser um movimento plural. São dois joelhos estudados ao mesmo tempo e, em diversos instantes há sobreposição de ambos, o que também faria com que o algoritmo de rastreamento perdesse a referência.

A Figura 17 ilustra a metodologia criada para a análise de movimentos de dobra de joelhos.



Figura 17: Diagrama de blocos do método de análise de movimentos de dobra dos joelhos.

Primeiramente, a fim de diminuir a quantidade de informação a ser processada, delimitou-se a região dos vídeos onde se encontram os joelhos dos músicos. Para efeito de análise, utilizou-se apenas a metade inferior dos vídeos.

Assim, pôde-se focar apenas no objetivo, eliminando informações excedentes. Como não há grandes mudanças de distância entre os músicos e a câmera – o instrumentista sempre está perfeitamente enquadrado na cena – é possível inferir que os joelhos sempre estarão localizados na metade inferior da filmagem, na direção vertical do tamanho do quadro. Por isso foi possível simplificar a análise eliminando os dados da metade superior.

Delimitada a região, decidiu-se então pelo uso da técnica de projeção de perfil horizontal (KOERICH, 1997; ZRAMDINI; INGOLD, 1993). Esta técnica é uma contagem de *pixels* partindo de um referencial inicial até um alvo. Para este caso em si, o referencial é a própria margem esquerda do quadro. Já o alvo, evidentemente, é o músico. Fazendo a contagem de *pixels* neste intervalo, gera-se um gráfico do estilo barras horizontais, em uma relação $cp^{q,h}$, em pixels, onde “q” é quadro do vídeo em análise e “h” é a linha deste quadro que gerou tal quantidade de *pixels*, criando o perfil, a silhueta do alvo proposto. Como “h” é relativo às linhas do vídeo, para este caso “h” varia de $L/2$ a L , onde L é a quantidade de linhas do quadro, ou seja, uma varredura da metade inferior do quadro.

Para a atual base de dados – devido ao pré-processamento realizado –, a contagem dos pixels pôde ser simplificada para a mera localização dos primeiros *pixels* nível 0 em cada linha que compõe a imagem. Graças à limiarização bimodal aplicada, assume-se que onde há *pixels* 0, está presente o músico. Portanto, localizando o primeiro *pixel* preto de cada linha e exibindo este valor em gráfico de colunas, obtém-se o perfil projetado do músico. A Figura 18 ilustra o método.

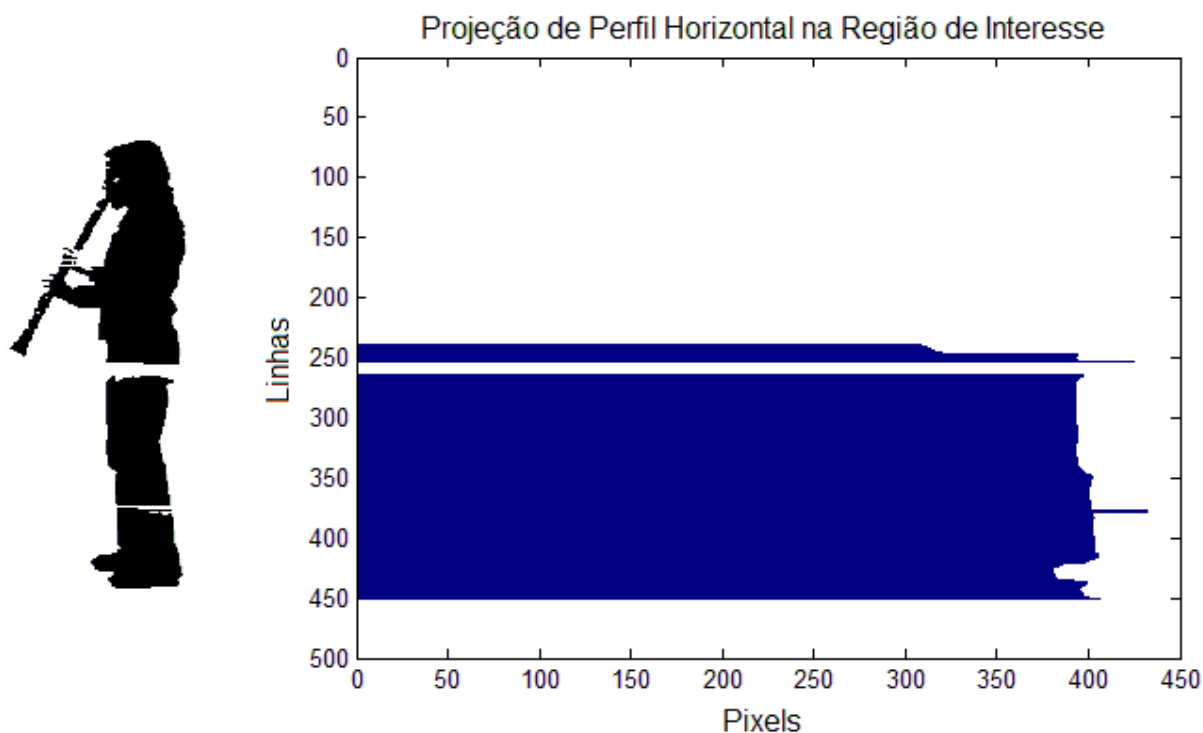


Figura 18: Projeção de perfil horizontal.

A proposta para detecção de gestos de dobra de joelhos é identificar alterações do perfil horizontal na região específica. Quando se dobra o joelho, a região tende a se destacar em relação ao restante da perna do músico. Os *pixels* pretos que representam o joelho tendem a deslocar-se para frente, reduzindo o valor no gráfico da projeção, como mostra a Figura 19. Observar esses instantes resultaria na identificação do movimento.

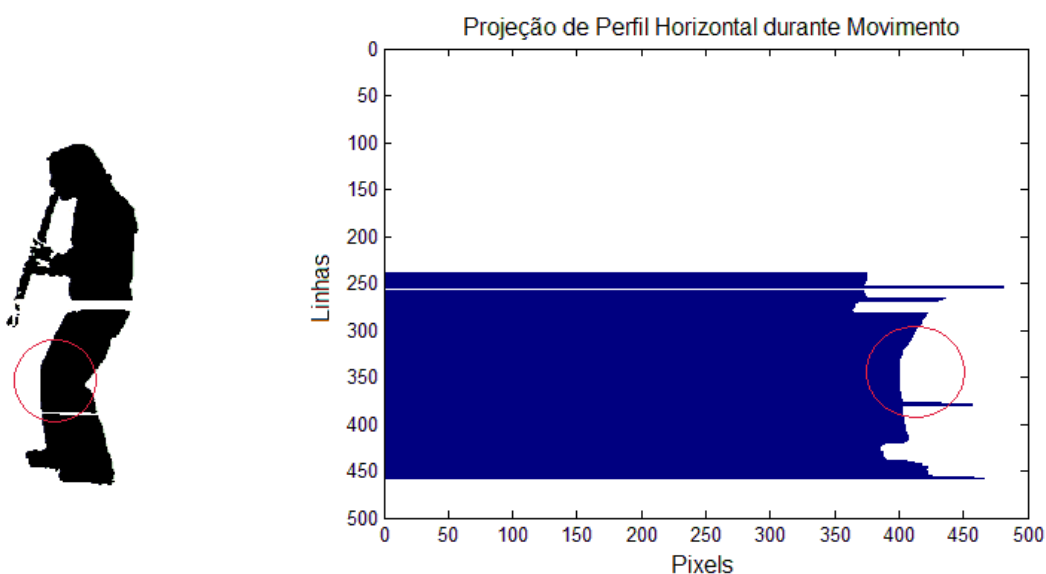


Figura 19: Projeção de perfil com dobra de joelhos.

O perfil da Figura 18, uniforme e vertical, agora se tornou quase que uma curvatura. Reparando na região destacada na Figura 19, nota-se que o principal responsável por esta mudança de comportamento são os joelhos. Portanto, para detectar os movimentos, bastaria criar um referencial fixo. Sempre que ele fosse atingido, indicaria o gesto.

Análogo ao estudo do clarinete, o referencial também precisa se manter coerente e não sofrer interferência de outros movimentos executados no decorrer do vídeo. Por isso, voltou-se a utilizar o centro de massa. No entanto, usar as coordenadas propriamente ditas não resultaria em detecção, uma vez que os joelhos sempre se encontram à frente do centro, ou, no máximo, na mesma posição. Dessa forma, foi necessário criar um referencial proporcional ao centro de massa (L_D), conforme ilustra a Figura 20.

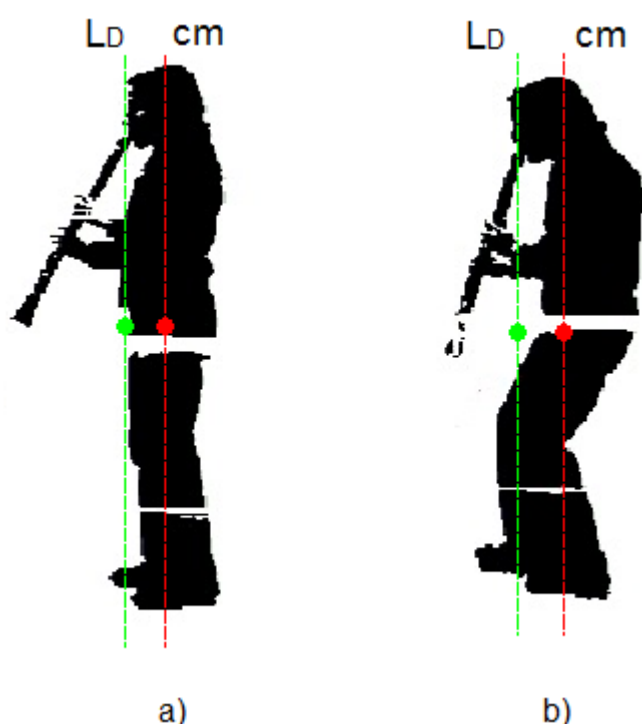


Figura 20: Referencial proporcional ao centro de massa.

Pelo fato de o centro de massa nada mais ser um ponto no quadro dos vídeos, o limiar de detecção (L_D) pôde ser constituído como um percentual da coordenada largura deste ponto. Assim, sempre que o clarinetista executasse o movimento (Figura 20a para Figura 20b), a contagem de pixels na região dos joelhos seria inferior a este referencial, possibilitando a detecção.

Agora se tem o seguinte quadro: contagem de pixels superior ao limite L_D , ou seja, não indicação de movimento, transformaria o atual valor da contagem para 0, ao passo que valores inferiores, extrapolando o limite, seriam alterados para 1. Esta operação possibilita a tomada de decisões: grupo de pixels 1 na região representam quadros em que os joelhos ultrapassaram o limiar, ou seja, produziram o movimento, ao passo que pixels 0 representam a não ultrapassagem do limiar, não caracterizando movimento.

Primeiramente optou-se por utilizar $L_D = 0,9 \cdot l_{CM}$, onde l_{CM} é a componente largura da coordenada do centro de massa do quadro em análise. Portanto, se a coordenada largura do centro de massa em um determinado quadro fosse, por exemplo, $l_{CM} = 340$, ter-se-ia $L_D = 306$ e o movimento seria detectado sempre que a contagem de pixels $cp^{q,[L/2,L]}$ fosse inferior a este número.

Assim como na análise dos movimentos de clarinete, o fator limitador para detecção de movimentos de joelho é específico para cada vídeo. No entanto, aparentemente a forma de atuação do clarinetista não influenciou na constituição do referencial, ou seja, cada clarinetista teve sua análise feita através do mesmo fator. Os valores obtidos para L_D encontram-se na Tabela 3.

Tabela 3: Média e desvio padrão da coordenada largura do centro de massa e fator limitador.

VÍDEO	cm_M	cm_{DP}	Fator	L_D
1.1	214	10	0.88	188
1.2	218	3	0.88	192
1.3	216	9	0.88	190
1.4	385	47	0.88	339
2.1	221	3	0.90	199
2.2	216	28	0.90	194
2.3	189	3	0.90	170
3	430	27	0.92	396
4	345	45	0.87	300

Uma vez definidos os valores de L_D , partiu-se para a limiarização das projeções de perfil horizontal, obtendo um resultado similar ao exibido na Figura 21.

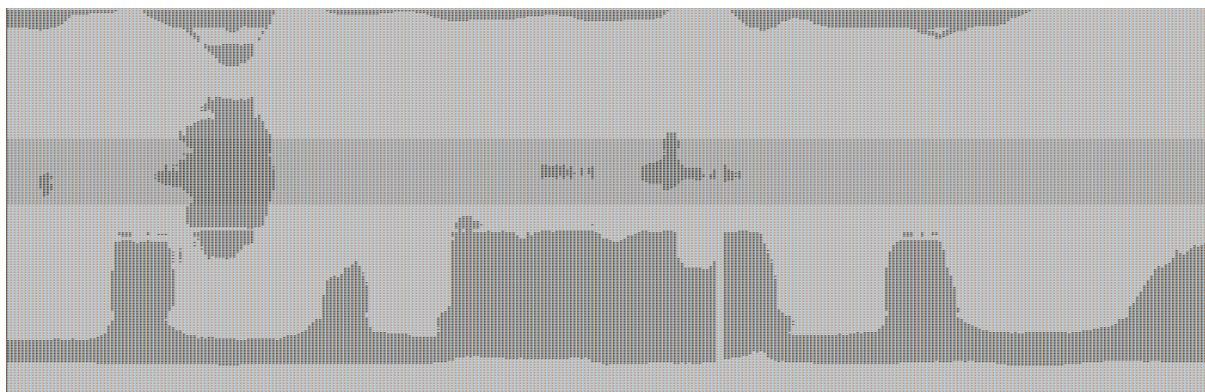


Figura 21: Exemplo de limiarização da projeção de perfil horizontal.

A Figura 21 é uma grande matriz de resultados da limiarização da projeção de perfil, onde as linhas desta matriz são equivalentes às próprias linhas do vídeo, enquanto que as colunas são os quadros. Por exemplo, ao se observar a célula (150,200) da matriz exemplificada na Figura 21, está sendo visualizado o resultado da limiarização da projeção de perfil gerada na linha 150 do quadro 200.

As áreas claras desta matriz indicam contagens superiores ao limiar, ou seja, não há movimento. A faixa central, em um tom intermediário de cinza, é o bloco de linhas da matriz que especificam a localização dos joelhos na imagem. Já as regiões mais escuras são resultados da limiarização inferiores a L_D , portanto, onde houve dobra de joelhos. As manchas escuras encontradas na borda superior e inferior da matriz representam respectivamente a ponta do clarinete e os pés dos músicos. Nestes dois casos a contagem de pixels sempre será inferior ao valor do limiar de detecção, gerando o aspecto escuro da Figura 21, mas não são importantes para o estudo.

A Figura 22 exemplifica o método. Observando-se a Figura 22a, nota-se que não há execução do movimento de dobra de joelhos. Logo, o referencial estipulado na Tabela 3 não foi ultrapassado, a contagem de pixels na região não foi inferior a L_D . Portanto, manteve-se o aspecto claro na faixa central da matriz (Figura 22b). Contudo, no instante mostrado na Figura 22c há execução do movimento. Ou seja, agora o referencial foi ultrapassado, a contagem de pixels foi inferior a L_D . Dessa forma, a porção central da matriz de resultados tornou-se escura (Figura 22d).

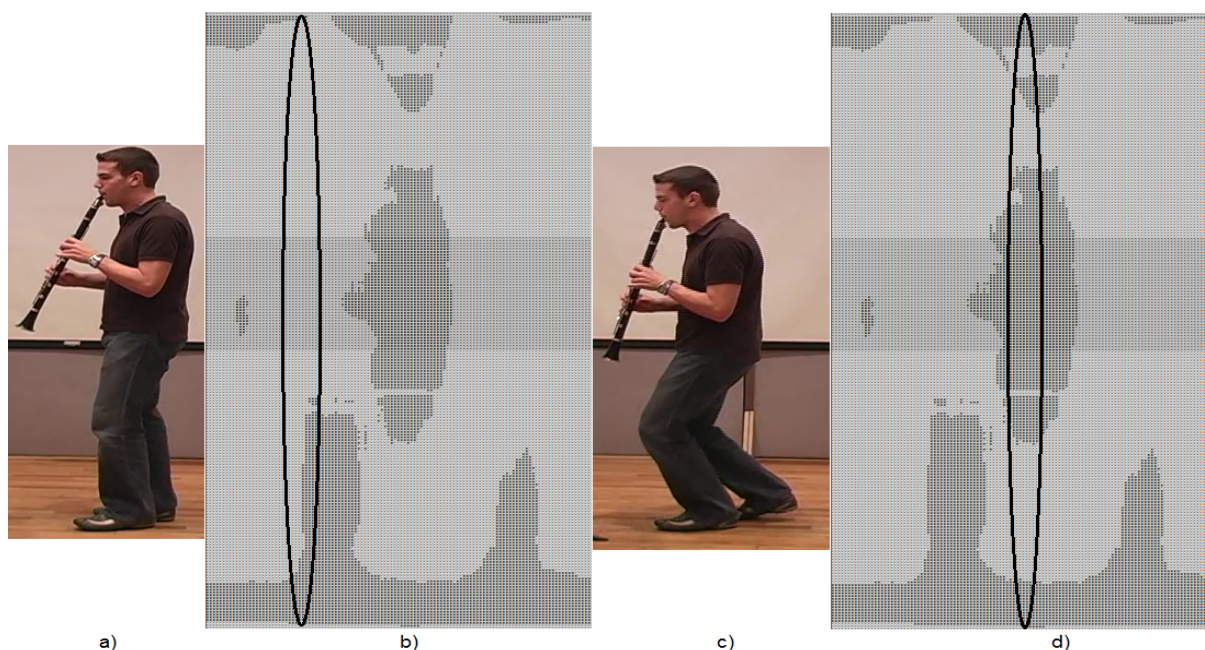


Figura 22: Comparação entre imagem e dados pós-processados.

As regiões da matriz dentro da faixa central que apresentam aspecto escuro representam o movimento. Manchas que englobam poucas colunas da matriz indicam movimentos de curta duração (poucos quadros), enquanto que manchas compreendidas entre várias colunas indicam que o movimento foi de grande duração. Quando o movimento é intenso, há a tendência de a região escura cobrir uma parte maior não só de quadros, mas de linhas do perfil, pois porções da perna que compõem a região do joelho também ultrapassam o limitador.

Os resultados obtidos nos nove vídeos serão exibidos e comentados na Seção 4.

3.5.3 Movimentos de Ondulações nas Costas

Assim como no estudo do movimento de joelhos, a análise dos gestos auxiliares de ondulações nas costas partiu da combinação de duas estratégias: determinação da região de interesse e uso da projeção de perfil. Novamente não se pôde utilizar o rastreamento de pontos pelo simples fato de não haver informação disponível para tal. A manutenção do ponto-referência numa região constante das costas é quase que impraticável. A Figura 23 ilustra o método proposto.

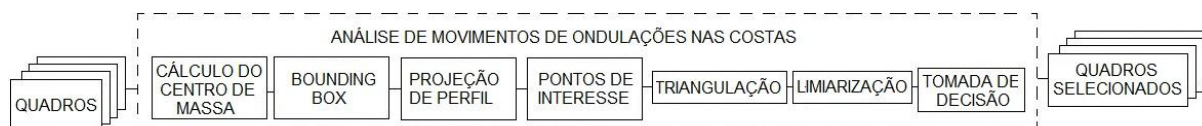


Figura 23: Diagrama de blocos do método de análise de movimentos de ondulações nas costas.

A projeção de perfil para o caso dos joelhos era mais simples. Como se tratava de uma projeção horizontal, varrendo quase que a totalidade da resolução do quadro, o perfil resultante certamente abrangia a região de interesse – os joelhos. Entretanto, para o estudo do movimento das costas, isto não é mais verdade. Como o movimento das costas é um movimento muito mais amplo, não apenas uma pequena porção do perfil sofrendo variação, detectar mudanças na formação do perfil horizontal agora aplicado da margem direita até o músico não resultaria em informações conclusivas.

O movimento de ondulações nas costas é de certa forma muito parecido com o movimento pendular de clarinete, ou seja, em essência, um translado vertical. Portanto, para tal estudo, fez-se ideal a aplicação do conceito de contagem de pixels também na vertical, ou $cp^{q,v}$, com “q” sendo o quadro em análise e “v” as colunas do vídeo.

Contudo, simplesmente gerar o perfil vertical geraria muitos dados desnecessários. Como o instrumentista fica na região mediana do vídeo, praticamente dois terços dos dados gerados seriam contagens tendendo ao tamanho do quadro, não contribuindo para a análise. Por outro lado, não se tem como restringir esta mesma contagem a uma porção central do vídeo, assim como fora feito no caso dos joelhos, porque o músico se desloca pelo cenário durante a execução, o que culminaria com a perda de informação. Por haver essa dificuldade para obtenção de dados, é que se decidiu pelo uso do algoritmo de *bounding box* (BRADSKI; KAEHLER, 2008).

Este contorno ajustável no instrumentista serve como referencial para com certeza garantir a sua presença. É como se a câmera focasse apenas no clarinetista, alterando sua resolução de filmagem conforme necessário, garantindo apenas o enquadramento do músico e seu instrumento. A partir disso, aperfeiçoaria a projeção de perfil vertical, que apenas atuaria no músico, sem perdas ou excesso de informação. A Figura 24a exemplifica uma *bounding box* utilizada no trabalho.

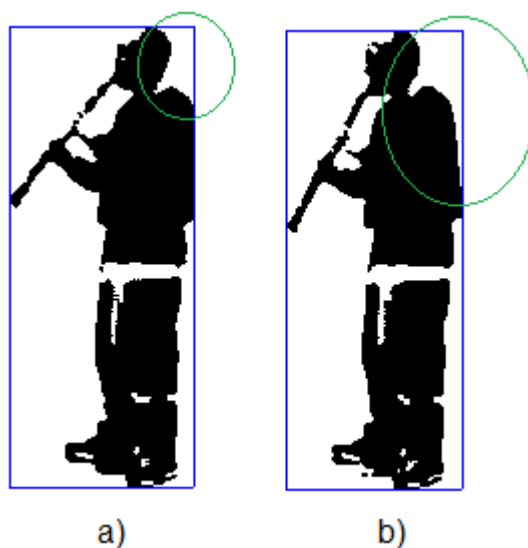


Figura 24: Uso da *bounding box*.

Projetar o perfil agora interno a *bounding box* garante o foco apenas no músico. Agora se tem a limitação $B0 \leq v \leq B1$, sendo o intervalo $[B0, B1]$ as extremidades laterais da *bounding box*. Sempre que houvesse um movimento de ondulação nas costas, a cabeça do músico iria para frente, alterando $B0$, mantendo a outra borda – $B1$ –, aumentando a área delimitada por esta borda e a cabeça, conforme a Figura 24b. Esta mudança de área indicaria uma mudança no perfil projetado na mesma, possibilitando a caracterização do movimento.

Isso tudo é bem intuitivo, porém ainda implica em um problema: como a *bounding box* limita todo o corpo do instrumentista, instantes de inclinação do corpo, ou de deslocamento pelo cenário, poderiam alterar o limitador, ou, a borda direita. Os pés do instrumentista poderiam assumir posições de destaque na formação da caixa, interferindo diretamente na projeção do perfil, causando falsas detecções de movimentos, como ilustra a Figura 25a. Pensando nisso que se restringiu a *bounding box* até apenas o centro de massa, conforme a Figura 25b.

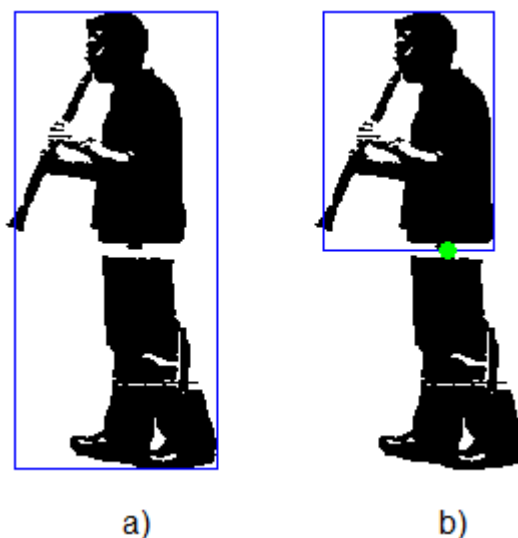


Figura 25: Comparação entre *bounding box* completa e restrita.

Aplicada a BB nos vídeos, pôde-se inserir a contagem de pixels vertical (KOERICH, 1997; ZRAMDINI; INGOLD, 1993) apenas na região delimitada pela caixa. Como o centro de massa sempre se localiza na cintura do músico, toda possível interferência dos pés na formação do contorno é neutralizada. Agora sim se tem a projeção de perfil vertical adequada para análise, como ilustrado na Figura 26.

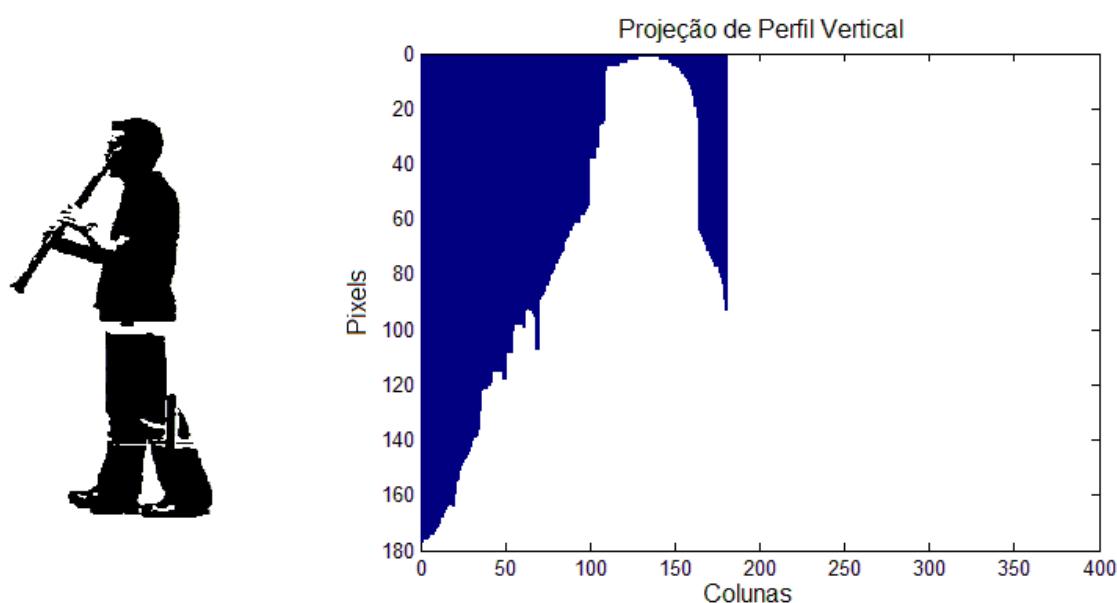


Figura 26: Projeção de perfil vertical apenas na região de interesse.

Ao se observar a Figura 26, é possível reparar na área que indica o movimento, limitada pela borda direita do contorno e pela cabeça do músico. Já a Figura 27 ilustra o aumento aparente da área em um dado momento de ocorrência

de movimento.

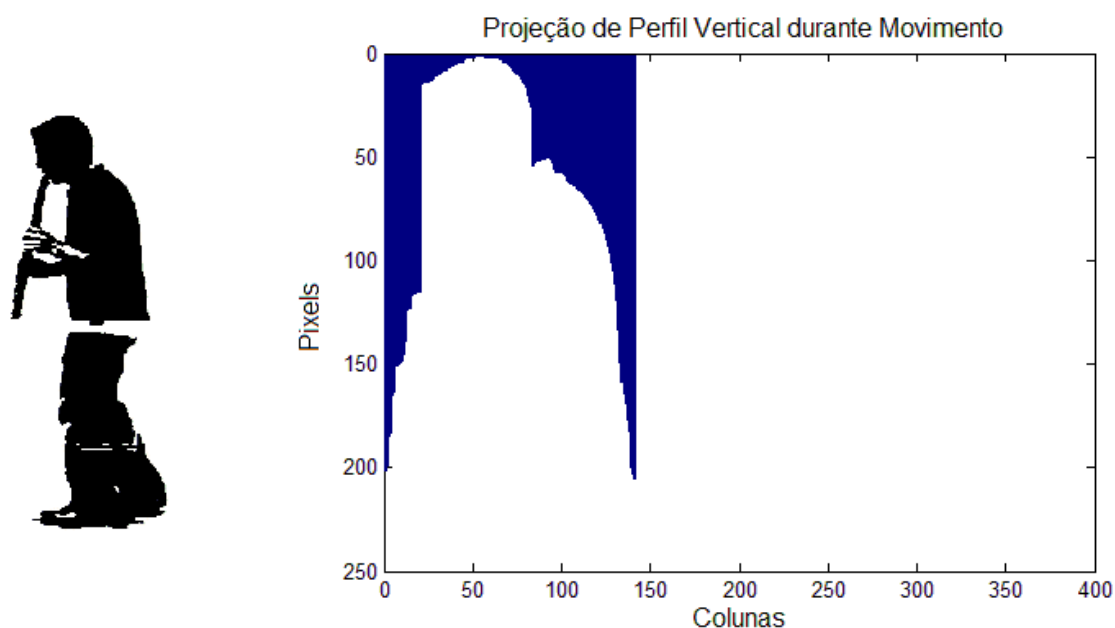


Figura 27: Projeção de perfil vertical apenas na ocorrência de movimento.

Comparando a Figura 26 e a Figura 27, nota-se o claro aumento da região de interesse. Baseado nesta prerrogativa, é que se passou a determinar uma forma de medir esta variação.

A variação em si ocorre em basicamente dois elementos: quantidade de elementos no perfil e o pico máximo dele. Uma vez que as costas não apresentam deformações, e ou saliências, o pico máximo sempre ocorrerá na borda criada pela *bounding box*, ao passo que quanto maior a inclinação da cabeça, haverá mais elementos no perfil. Estes dois parâmetros foram denominados “pontos de interesse” e passaram a ser determinantes no estudo dos gestos.

O algoritmo desenvolvido passou a procurar dentro da projeção de perfil vertical um ponto p1 e um ponto p2, formados por uma coordenada relativa ao número de colunas do perfil vertical e outra relativa à quantia de *pixels* desta coluna. O ponto p1 é o primeiro ponto partindo da borda esquerda da *bounding box* até atingir a cabeça do músico e o ponto p2 é a última coluna formadora da BB.

Observando os perfis, pôde-se notar que a cabeça dos instrumentistas sempre era encontrada em contagens de *pixels* inferiores a 15. Dessa forma, arbitrariamente optou-se por determinar o ponto p1 como sendo a primeira coluna da projeção de perfil, partindo da origem, que apresentasse contagem inferior a 15

pixels. Já p2 seria a última coluna formadora do perfil. A obtenção dos pontos está demonstrada na Figura 28.

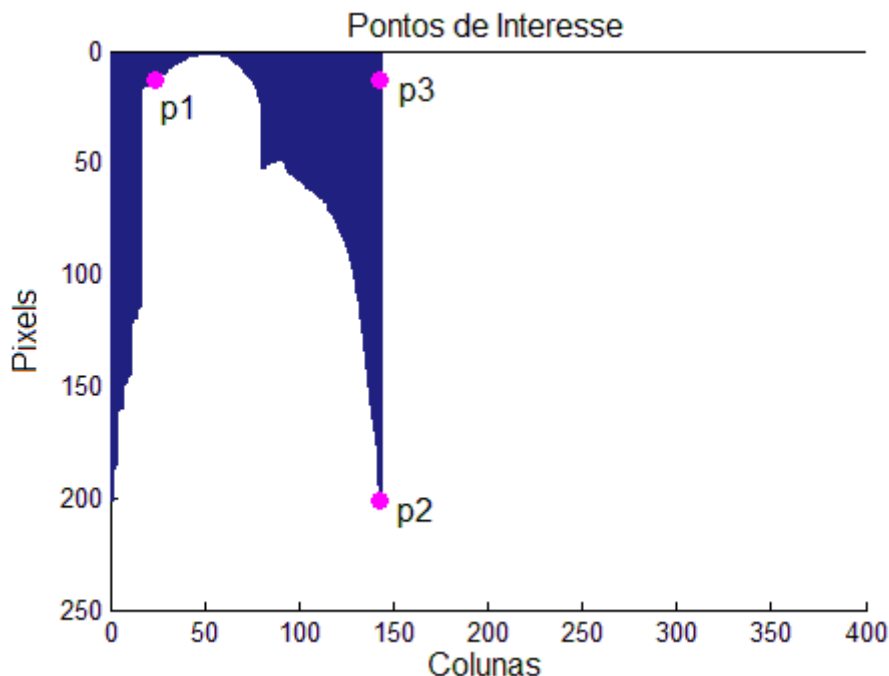


Figura 28: Pontos de interesse.

É importante ressaltar que os pontos de interesse não possuem relação alguma com o rastreamento de marcadores presos aos músicos. Os pontos de interesse são coordenadas importantes retiradas diretamente do perfil calculado. Não há rastreamento, mas sim uma análise de quantidade de pixels.

Na Figura 28 é possível perceber um terceiro ponto, p3, que até agora não foi mencionado. Este ponto nada mais é que um referencial para criação de um triângulo retângulo entre os dois pontos de interesse p1 e p2. Por isso que o método de análise de movimentos das costas foi denominado "triangulação". Analisar a variação da área do triângulo formado pelos pontos trará um resultado proporcional à variação da área apenas da região do perfil, possibilitando a identificação do movimento. O método proposto está ilustrado na Figura 29.

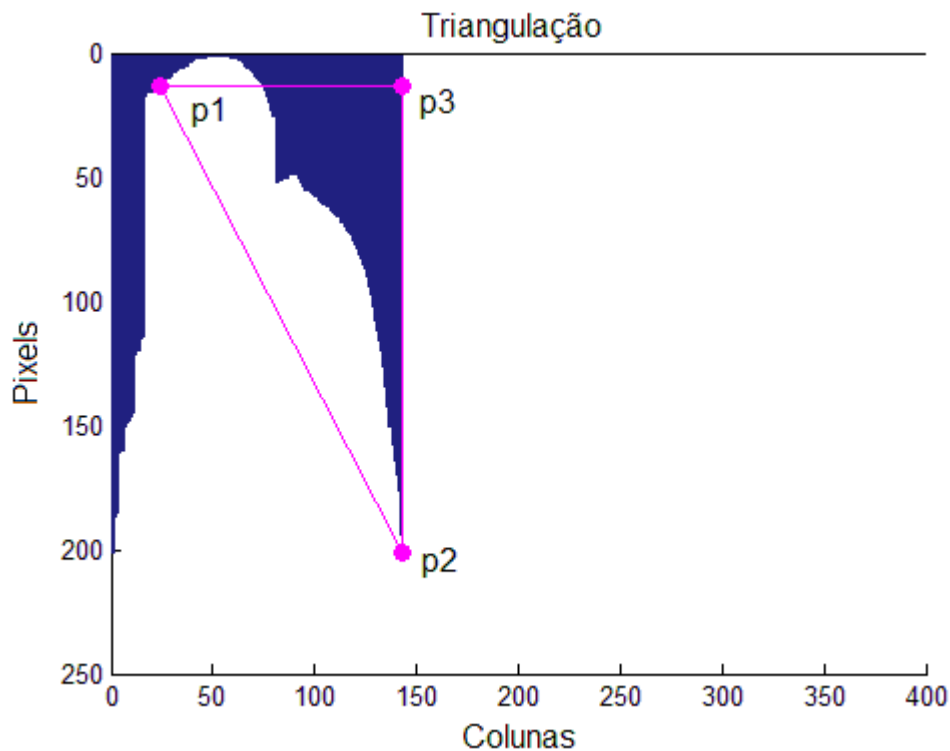


Figura 29: Triangulação.

O termo “área” pode parecer errôneo, pois não se tratam de medidas descritivas. A área aqui nada mais é que a relação direta entre quantidade de elementos no perfil – a base do triângulo – e o pico máximo – a altura do triângulo.

A partir dos conceitos de geometria analítica (STEINBRUCH; WINTERLE, 1987), sabe-se que a área (S^q) compreendida entre os pontos p1, p2 e p3, em um quadro “q”, pode ser calculada através equação 2 a seguir:

$$S^q = \frac{\overline{p1,p3} \times \overline{p2,p3}}{2} \quad (2)$$

Dessa forma, o parâmetro S^q passa a ser o elemento avaliador de movimentos, em ordem de grandeza elementos x *pixels*.

Agora, valores grandes de S representam ondulações nas costas, ao passo que pequenos valores de S indicam repouso, na vertical. A Figura 30 ilustra a caracterização do movimento baseado na variação do fator área.

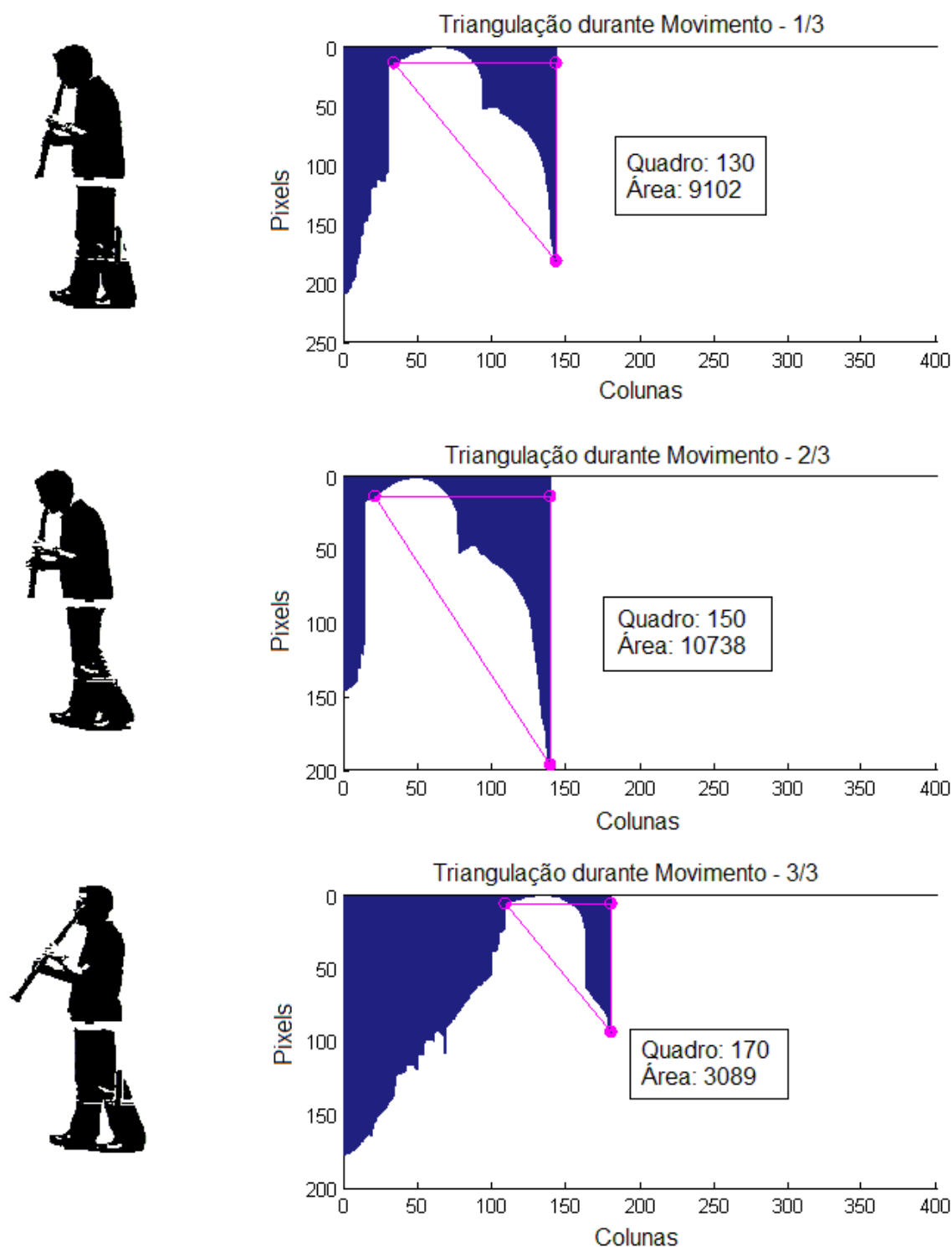


Figura 30: Variação da triangulação durante a presença de movimento.

Na Figura 30, pode-se notar que na transição de 40 quadros (aproximadamente 1,5 segundos) ocorre um movimento de costas. A área proposta pelo método de triangulação parte de um valor 9000 para 11000 e reduz para 3000. Repara-se ainda que caso a *bounding box* não fosse limitada à cintura do músico, os

pés aumentariam a área do contorno, interferindo diretamente na produção do perfil e no cálculo da área.

Gerados os pontos p_1 , p_2 e p_3 , assim como os valores de S^q para todos os quadros da base de vídeos, bastou-se aplicar a tomada de decisão baseada em um limiar. Assim como também sugerido no estudo de movimentos de clarinete, o limiar L foi criado a partir das médias (S_M) e desvios padrão (S_{DP}) das áreas obtidas. A Tabela 4 mostra os valores então obtidos.

Tabela 4: Média e desvio padrão do método de triangulação.

VÍDEO	S_M	S_{DP}	$L (S_M + S_{DP})$
1.1	1461	830	2291
1.2	1326	569	1895
1.3	1273	830	2103
1.4	4580	2107	6687
2.1	997	224	1221
2.2	1618	743	2361
2.3	1436	505	1941
3	3510	1798	5308
4	6101	3017	9118

Os nove vídeos foram avaliados segundo o método proposto e os respectivos resultados serão exibidos e comentados na Seção 4.

Com toda a gama de informações geradas pelos algoritmos anteriormente descritos, pode-se partir para a etapa final: a avaliação dos resultados gerados pelos algoritmos que compõem o método proposto.

A Figura 31 apresenta todos os elementos do método proposto para detecção dos três movimentos auxiliares propostos em um único bloco.

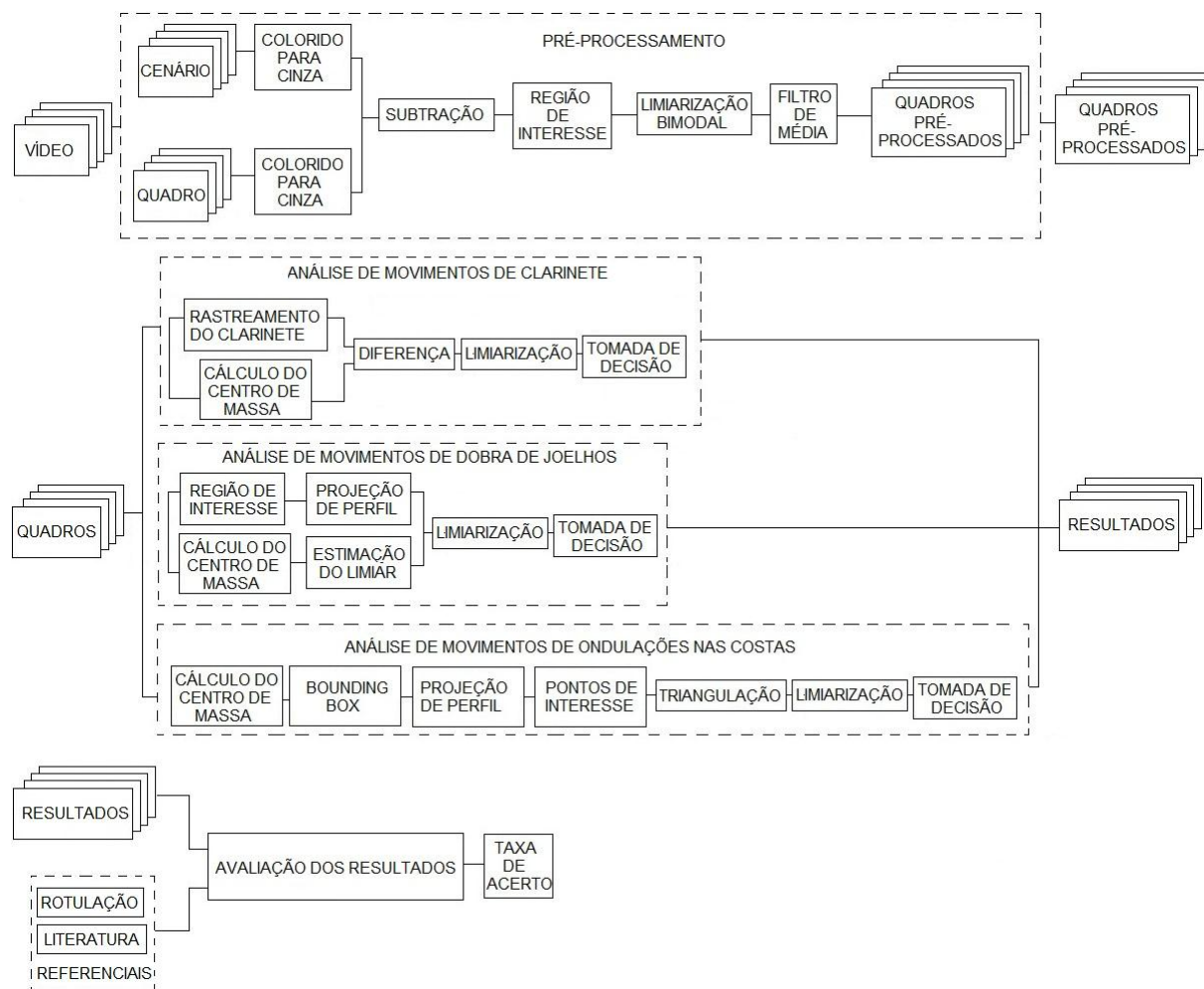


Figura 31: Diagrama final da metodologia.

4 AVALIAÇÃO DOS RESULTADOS

A partir de agora, cada seção do método proposto no Capítulo 3 terá seus resultados apresentados e comentados.

4.1 ANÁLISE DE MOVIMENTOS DE CLARINETE

Aplicado o método proposto na Seção 3.5.1, obteve-se um resultado em forma de tabela contendo os intervalos de quadros de detecção de movimentos e uma representação gráfica. Para ilustrar, toma-se por base o resultado obtido para um dos vídeos do clarinetista 1, que está apresentado na Figura 32 e na Tabela 5 a seguir.

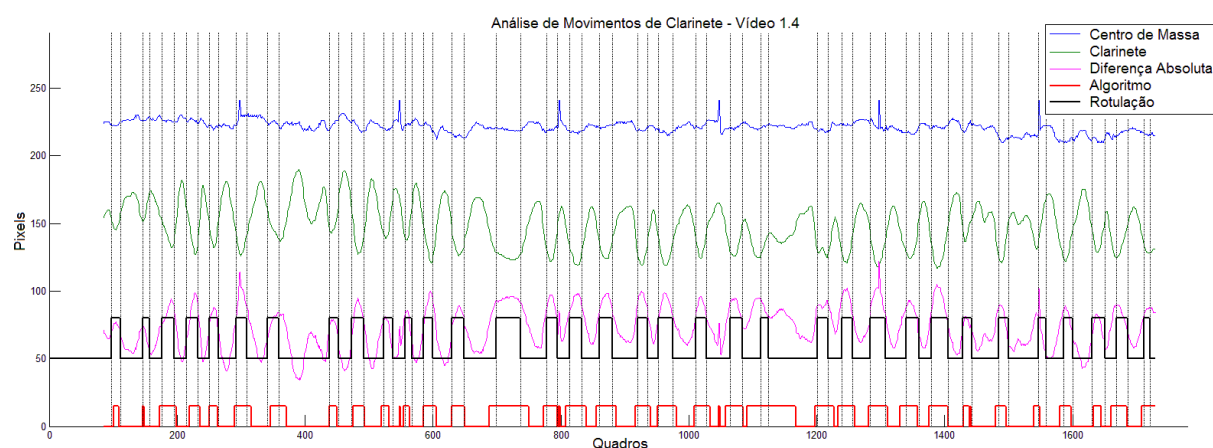


Figura 32: Resultado da análise do movimento de clarinete para o vídeo 1.4.

Tabela 5: Resultado da análise do movimento de clarinete para o vídeo 1.4.

GESTO	Intervalo de Quadros		RESULTADO
	ROTULAÇÃO	ALGORITMO	
1	97 - 112	102 - 109	PV
2	147 - 157	147 - 149	PV
3	176 - 195	174 - 199	PV
4	215 - 233	220 - 236	PV
5	251 - 265	252 - 264	PV
6	292 - 309	291 - 316	PV
7	341 - 359	347 - 371	PV
8	438 - 452	439 - 450	PV
9	473 - 492	476 - 493	PV
10	523 - 537	521 - 532	PV
11		549	
12	556 - 567	555 - 564	PV
13	585 - 600	586 - 605	PV
14	630 - 648	631 - 650	PV
15	699 - 737	689 - 750	PV
16	778 - 795	774 - 795	PV
17		798 - 799	
18	813 - 833	809 - 840	PV
19	860 - 880	857 - 887	PV
20	920 - 935	917 - 940	PV
21	952 - 975	952 - 981	PV
22	1011 - 1027	1010 - 1033	PV
23		1048 - 1049	
24	1065 - 1084	1059 - 1085	PV
25	1113 - 1125	1092 - 1167	PV
26	1201 - 1218	1198 - 1227	PV
27	1239 - 1256	1234 - 1260	PV
28	1284 - 1307	1282 - 1311	PV
29	1340 - 1360	1331 - 1358	PV
30	1379 - 1405	1377 - 1405	PV
31	1428 - 1442	1430 - 1439	PV
32		1441	
33	1484 - 1500	1480 - 1496	PV
34	1546 - 1558	1540 - 1549	PV
35	1586 - 1600	1581 - 1598	PV
36	1630 - 1651	1634 - 1645	PV
37	1668 - 1686	1661 - 1685	PV
38	1711 - 1720	1709 - 1730	PV

A Tabela 5 apresenta o resultado obtido a partir do algoritmo proposto, comparado com a rotulação prévia. PV é a representação de detecções corretas, positivos verdadeiros, FP são as falsas detecções, falsos positivos e FN são falsas não detecções, falsos negativos.

É possível perceber que a rotulação e o algoritmo não geram resultados idênticos. Há intervalos de quadros maiores e outros menores. Existe ainda uma defasagem de tempo entre a detecção via observação e a computacional. Tudo isso é aceitável devido à rotulação não ser exata. O algoritmo indica o movimento toda vez que o limiar L da curva de diferenças for atingido. Se este não for alcançado, não há detecção. Já a rotulação é subjetiva, dependente do observador. O reconhecimento do movimento não é automático, depende de quando que o observador foi sensibilizado pelo gesto, seja no seu início, na sua duração e no seu fim. Isso justifica a defasagem de resultados.

Observando a Tabela 5, encontram-se resultados detectados pelo algoritmo, mas que não foram observados na rotulação. É o caso dos movimentos 11, 17, 23 e 32. Aparentemente são falsos, pois não foram observados na rotulação prévia, no entanto, numa segunda observação, foram percebidos, ou seja, são também de caráter PV. Este resultado é muito interessante porque elucida o real objetivo da pesquisa: substituir a observação dos vídeos. Neste caso, os 4 movimentos seriam ignorados nas futuras análises, representando perda de dados. O método de detecção possibilitou cobrir essas falhas.

Daqui por diante, as colunas de resultados das tabelas de análise não terão mais espaços em branco. Situações como as ocorridas no exemplo citado anteriormente já serão exibidas com a verificação dos resultados sugeridos pelo algoritmo. Dessa forma, onde não há rotulação prévia, mas o algoritmo sugeriu presença de movimento, seria FP caso a segunda análise determinasse que de fato não houvesse movimento, portanto um erro do algoritmo, ou PV caso realmente houvesse movimento.

Tomando como exemplo os movimentos 11, 17, 23 e 32 da Tabela 5, o algoritmo sugere movimentos com 2 quadros de duração. Segundo a Tabela 1 do Capítulo 3, os vídeos do clarinetista 1 contém 25 quadros por segundo, ou seja, esses gestos detectados teriam aproximadamente 80ms de duração. Isso não é plausível. O que ocorre na verdade é que L foi extrapolado durante apenas 2

quadros. Um limitador menor aumentaria este intervalo. Por isso pode-se inferir que o algoritmo não é robusto o suficiente para mensurar movimentos, seja em sua magnitude ou em duração, mas sim indicar instantes nos vídeos onde há movimentos.

Outro problema ocorrido nas análises é o que será chamado de “trens de movimentos”. Em regiões do gráfico onde há oscilação justamente na área de aplicação da limiarização, surgem diversos pulsos em sequencia, como pode ser visto nos movimentos 4, 9 e 13 da Tabela 6.

Tabela 6: Resultado da análise do movimento de clarinete para o vídeo 3.

GESTO	Intervalo de Quadros		RESULTADO
	ROTULAÇÃO	ALGORITMO	
1	347 - 376	366 - 377	PV
2	559 - 574	563 - 574	PV
3	666 - 681	671 - 673	PV
4	807 - 823	798-803;814-820	PV
5		1049	FP
6	1206 - 1221		FN
7		1253 - 1288	PV
8	1299 - 1322	1324 - 1345	PV
9		1382;1384-1425	PV
10		1548 - 1549	PV
11	1729 - 1745	1721 - 1737	PV
12		1801 - 1857	FP
13	1879 - 1917	1862-1897;1912-1924	PV
14	1982 - 2030	1978 - 2021	PV

Esses pulsos não significam diversos movimentos, mas uma falha de tomada de decisão como ilustra a Figura 33.

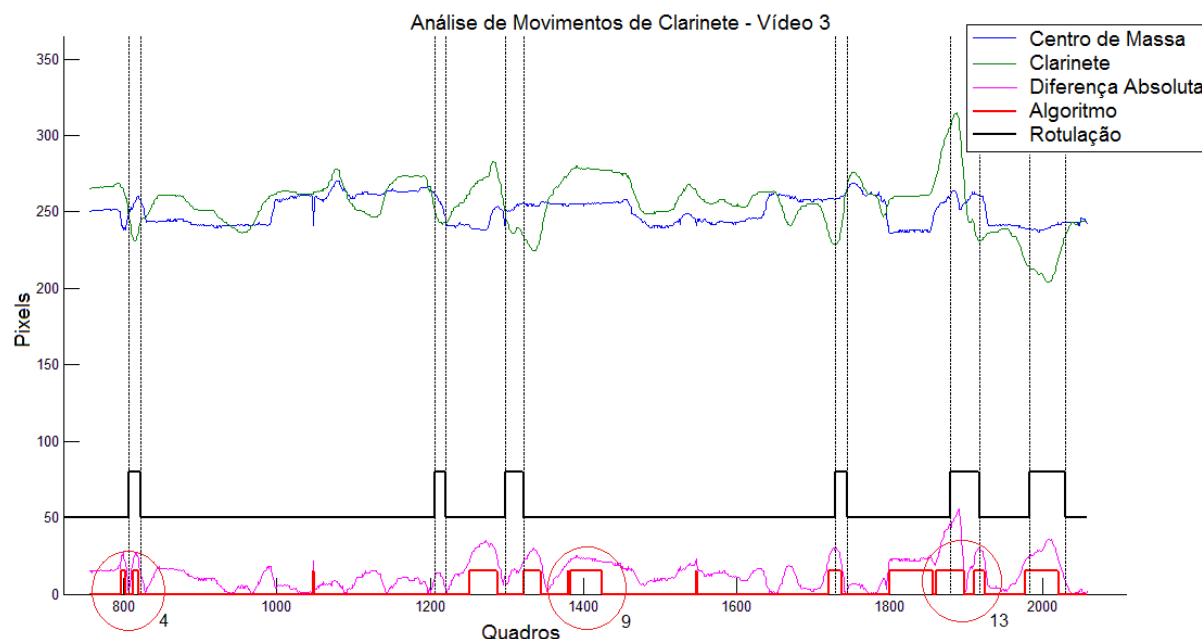


Figura 33: Efeito de trem de movimentos.

Os movimentos 4, 9 e 13 da Tabela 6, assim como as regiões 4, 9 e 13 destacadas na Figura 33 ilustram o efeito de trem de movimentos. Nestes três casos ocorreram movimentos muito intensos do clarinete. Iniciou em um ponto abaixo do centro de massa, indo para um ponto máximo e retornando para um valor mediano. A operação de diferença absoluta atingiu o limiar L quando o clarinete atingiu o valor abaixo do centro de massa e, ao passo que foi subindo, aproximou-se do referencial, não mais extrapolando L . No entanto, o movimento continuou crescente, atingindo novamente o limiar. Esta inflexão que gerou o efeito.

O efeito de trem de movimentos foi comum em praticamente todos os vídeos. Alguns, como neste caso, devido a *performance*, outros devido à trajetória do marcador virtual ser muito parecida com a do centro de massa em se tratando da componente vertical, ou seja, o limiarização ficou muito tênue: mínimas variações acionavam a detecção. Outros casos ainda apresentaram pequenas falhas de constituição das trajetórias do marcador virtual ou do cálculo do centro de massa, também culminando com tal efeito.

Com o intuito de simplificação de resultados, os trens de movimentos foram filtrados. Detecções que apresentaram esse padrão foram vistoriadas. Quando se percebeu que a curva de diferença mantinha certo padrão de comportamento, ou que na região ocorreu um espúrio nos dados que geraram a curva, optou-se por eliminar o efeito de pulsos causados pela limiarização.

Todos os resultados da análise, com seus respectivos gráficos, estão referenciados da Tabela 12 a Tabela 19 e da Figura 44 a Figura 51 do Apêndice. A Tabela 7 sintetiza os resultados da análise de movimentos auxiliares de clarinete obtidos sobre os nove vídeos da base de dados.

Tabela 7: Resultado final da análise de movimentos de clarinete.

ROTULAÇÃO	ALGORITMO					TAXA DE ACERTO
	POSITIVOS VERDADEIROS	FALSOS POSITIVOS	TOTAL DE DETECÇÕES	FALSOS NEGATIVOS	TOTAL DE GESTOS	
151	192	53	245	9	254	75,6%

Observando-se a Tabela 7, nota-se uma grande diferença entre a quantidade de movimentos detectados pela observação dos vídeos e os sugeridos pelo método de detecção – 94 detecções a mais. No entanto, mais da metade dessas detecções a mais foram de caráter falso positivo.

Comparados os 151 gestos auxiliares de clarinete rotulados com os 245 movimentos sugeridos pelo algoritmo, houve 9 movimentos rotulados, mas não observados pelo método proposto. Este resultado não é ideal, seriam 9 gestos descartados para as futuras análises, porém 9 gestos em meio a 192 representam apenas 5% de falhas de detecção. Além disso, o método encontrou 41 movimentos que a rotulação prévia não indicou. Como fora explicado, a observação humana é altamente sujeita a erros e isso fica evidenciado aqui.

Somando os 53 movimentos erroneamente detectados com as 9 falhas de detecção, totalizam-se 62 falsos resultados, em um universo de 254 gestos auxiliares de clarinete. Isso produz uma taxa de acerto de aproximadamente 75% para o método proposto.

Observando-se as tabelas do apêndice, nota-se que 43 dos 53 resultados de falso positivo aconteceram nos vídeos do clarinetista 2. Como mostrado na Figura 10d, os vídeos deste clarinetista tiveram o pior resultado de pré-processamento, principalmente devido à roupa utilizada por ele. A roupa clara dificultou a limiarização bimodal da imagem, comprometendo o destaque do músico em relação ao cenário durante o vídeo. Este resultado evidencia a importância do pré-processamento adequado à imagem. Os demais erros de detecção ocorreram pelo fator L de limiarização das curvas de diferença (centro de massa e marcador virtual).

4.2 ANÁLISE DE MOVIMENTOS DE DOBRA DE JOELHOS

Uma vez que todos os nove vídeos da base de dados foram pós-processados pelo método apresentado na Seção 3.5.2, partiu-se para a análise dos gráficos gerados, a identificação dos intervalos de quadros contendo movimentos e a verificação dos mesmos com a rotulação prévia. Para ilustrar, os resultados do clarinetista 3 encontram-se exemplificados na Tabela 8 e na Figura 34 a Figura 39.

Tabela 8: Resultado da análise do movimento de joelhos para o vídeo 3.

GESTO	Intervalo de Quadros		RESULTADO
	ROTULAÇÃO	ALGORITMO	
1	320 – 326	307 – 319	PV
2	453 – 460	456 – 463	PV
3	558 – 561	566 – 672	PV
4	754 – 762	747 – 790	PV
5		995 – 1066	FP
6	1074 – 1082	1074 – 1214	PV
7	1291 – 1297	1303 – 1463	PV
8	1542 – 1546	1520 – 1553	PV
9		1571 – 1590	FP
10		1640 – 1721	FP
11	1746 – 1755	1730 – 1788	PV
12	1865 – 1879	1859 – 1868	PV

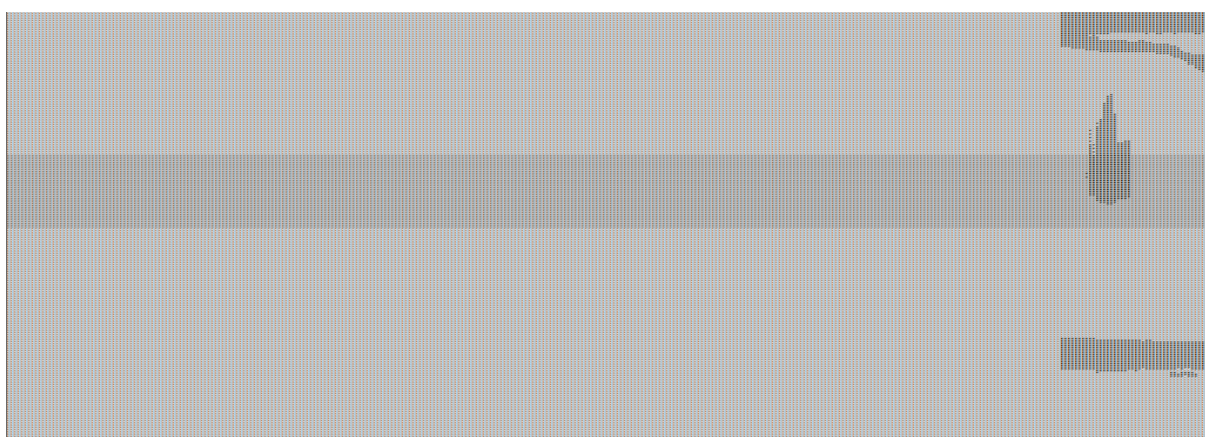


Figura 34: Resultado da análise do movimento de joelhos para o vídeo 3 (quadros 1 - 343).

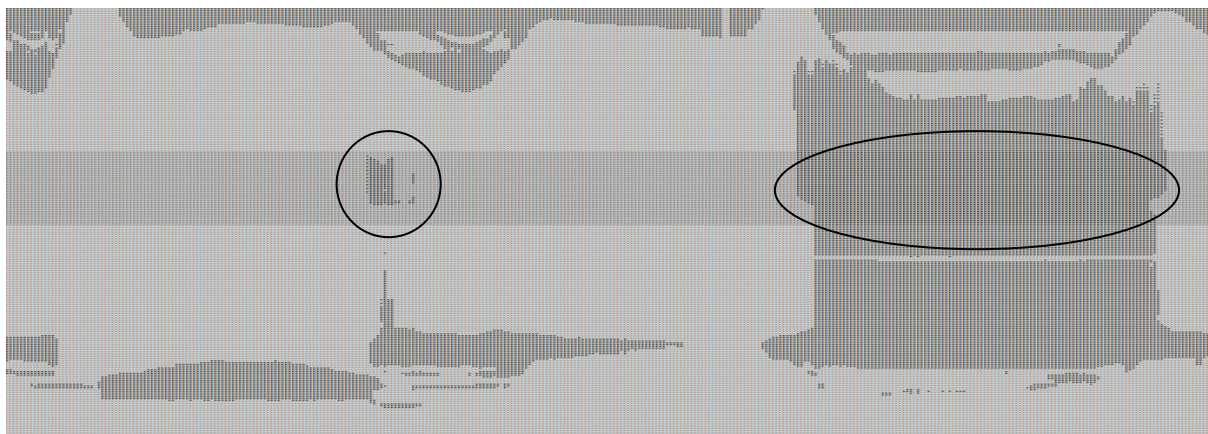


Figura 35: Resultado da análise do movimento de joelhos para o vídeo 3 (quadros 344 - 686).

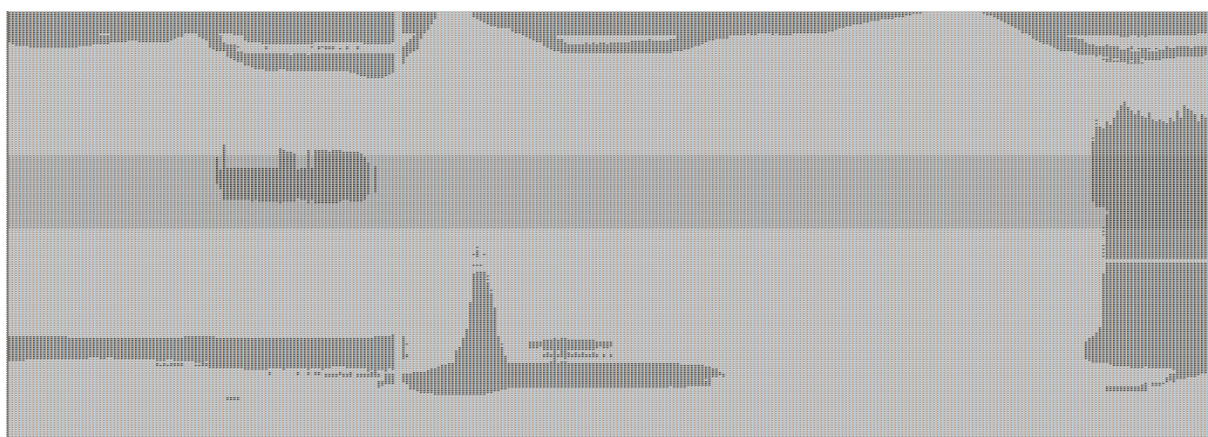


Figura 36: Resultado da análise do movimento de joelhos para o vídeo 3 (quadros 687 - 1029).

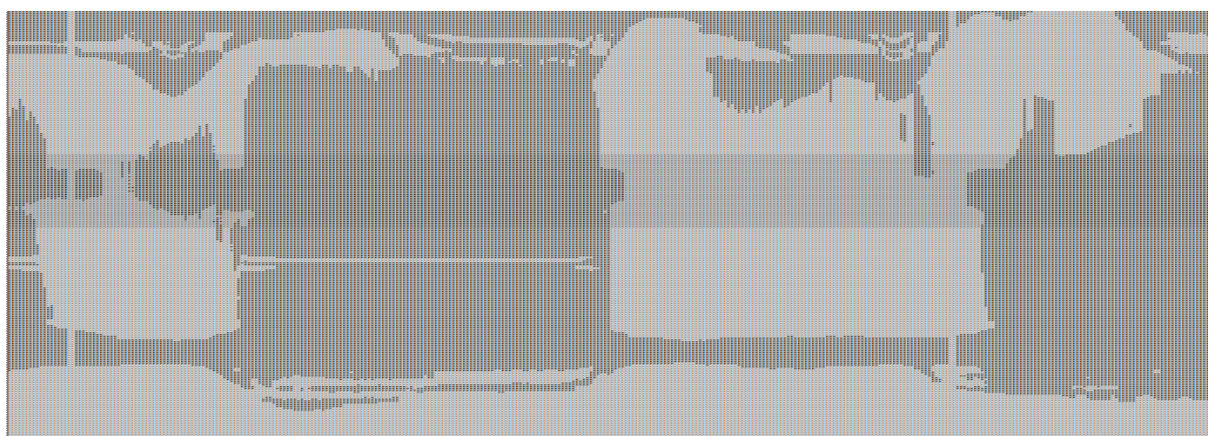


Figura 37: Resultado da análise do movimento de joelhos para o vídeo 3 (quadros 1030 - 1372).



Figura 38: Resultado da análise do movimento de joelhos para o vídeo 3 (quadros 1373 - 1715).

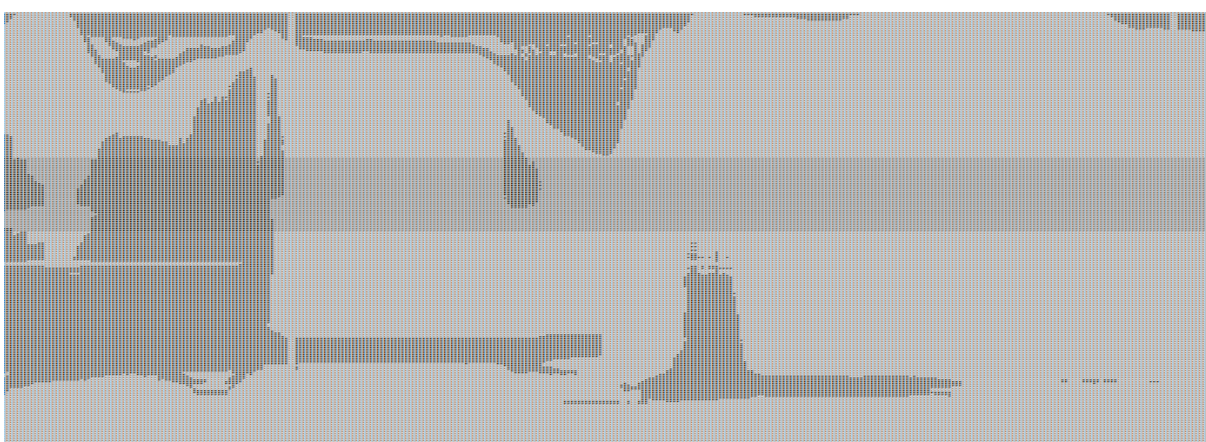


Figura 39: Resultado da análise do movimento de joelhos para o vídeo 3 (quadros 1716 - 2058).

Pode-se verificar que todos os movimentos identificados na observação, foram também identificados pela metodologia proposta. Três outros movimentos foram sugeridos, no entanto são de caráter falso positivo. Ocorreram em instantes que a instrumentista fez a abertura das pernas, em um deslocamento pelo cenário com sentido diferente das duas dimensões que o algoritmo é capaz de trabalhar, similar ao problema envolvido no deslocamento de peso, gerando confusão na tomada de decisão.

Aqui se pode explicar melhor a duração do movimento e sua respectiva representação gráfica. As marcações na Figura 35 ilustram os movimentos 2 e 3 da Tabela 8, respectivamente. Enquanto o movimento 2 é representado por uma pequena mancha escura na faixa central do gráfico, o movimento 3 é muito maior, seja no número de linhas ou no número de colunas. Como as linhas do gráfico representam as linhas que compõem o vídeo e as colunas do gráfico são os quadros, muitas linhas escuras indicam que uma maior porção do corpo ultrapassou o limiar de detecção L_D , ou seja, um movimento mais intenso, ao passo que muitas

colunas indicam um movimento mais demorado, os joelhos permaneceram dobrados por um maior tempo.

Ao se comparar este resultado com a detecção de movimentos de clarinete, nota-se que aqui é possível deduzir intensidade e duração de movimentos, enquanto que o mesmo não é válido para o clarinete. Já os problemas relativos a “trem de movimentos” não se fizeram presentes no estudo dos joelhos.

A Figura 36, a Figura 37, a Figura 38 e a Figura 39 completam a ilustração gráfica dos resultados obtidos na Tabela 8.

Todos os resultados da análise, com seus respectivos gráficos, estão referenciados da Tabela 20 a Tabela 24 e da Figura 52 a Figura 76 do Apêndice. A Tabela 9 sintetiza os resultados da análise de movimentos auxiliares de dobra de joelhos obtidos sobre cada um dos vídeos da base de dados.

Tabela 9: Resultado final da análise de movimentos de dobra de joelhos.

ROTULAÇÃO	ALGORITMO					TAXA DE ACERTO
	POSITIVOS VERDADEIROS	FALSOS POSITIVOS	TOTAL DE DETECÇÕES	FALSOS NEGATIVOS	TOTAL DE GESTOS	
54	58	17	75	10	85	68.2%

Observando a Tabela 9, nota-se que novamente o algoritmo detectou mais movimentos do que a rotulação prévia. Entretanto, analisando apenas os acertos de detecções, os números foram muito próximos.

Outra questão importante é a presença de 10 movimentos não detectados. Em se tratando de números absolutos, o erro foi quase que o mesmo da detecção de movimentos de clarinete (9 falhas), porém, as 10 não detecções de movimentos de dobra de joelhos representam cerca de 15% do total de gestos ocorridos, um número consideravelmente elevado.

Analisando os erros de detecção para os quatro instrumentistas, o principal causador de erros de detecção foi o posicionamento do músico em relação à partitura, ou à câmera. A análise dos vídeos só foi possível em duas dimensões, uma vez que não se tem gravações em 3D e nem se tem como obtê-las. Quando o clarinetista fica em um plano com uma angulação relativa à partitura, acaba produzindo movimentos na terceira dimensão, que só são observáveis no vídeo como uma projeção deles para as duas dimensões existentes.

Partindo desse pressuposto, quando o instrumentista apresenta uma angulação relativa à partitura, ou à câmera que está gerando o vídeo, nas duas dimensões do vídeo é como se ele estivesse de pernas abertas. Logo, quando há o deslocamento do peso de uma perna para a outra neste plano de gravação, na projeção analisada pelo algoritmo é como se tivesse ocorrido um movimento de dobra de joelho, uma vez que a região de interesse sofre um deslocamento instantâneo, sem praticamente alterar o centro de massa. 9 falsos positivos são justificáveis por este argumento. Um décimo falso positivo ocorreu por sobreposição dos joelhos os outros 7 falsos positivos, assim como 8 das 10 falhas de detecção ocorreram por falha de pré-processamento dos vídeos do clarinetista 2.

4.3 ANÁLISE DE MOVIMENTOS DE ONDULAÇÕES NAS COSTAS

Os nove vídeos da base de dados foram analisados pelo método apresentado na Seção 3.5.3. Como exemplo de resultado, a Figura 40 e a Tabela 10 mostram o comparativo entre rotulação, metodologia e verificação de resultados para o clarinetista 2.

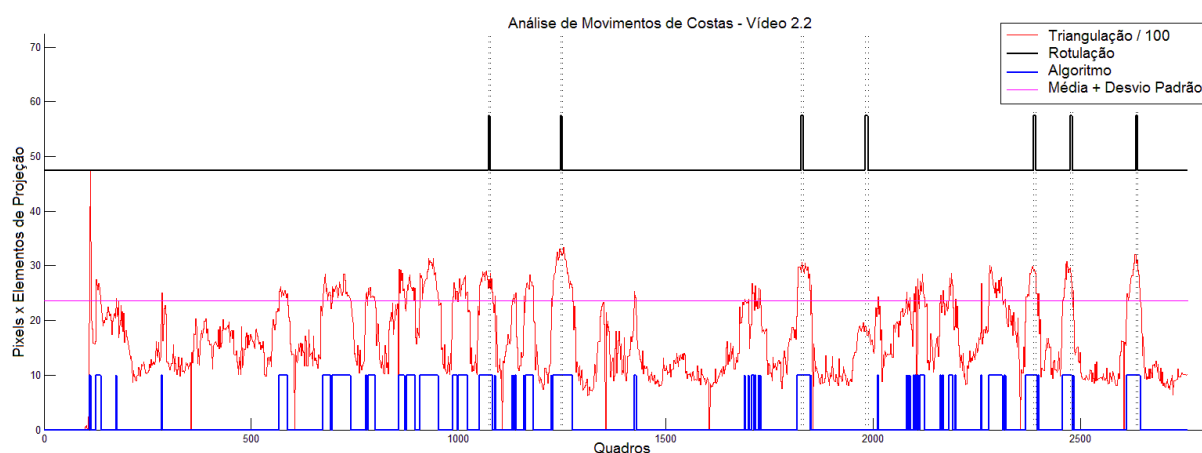


Figura 40: Resultado da análise do movimento de ondulação das costas para o vídeo 2.2.

Tabela 10: Resultado da análise do movimento de costas para o vídeo 2.2.

GESTO	Intervalo de Quadros		RESULTADO
	ROTULAÇÃO	ALGORITMO	
1		568 - 587	PV
2		672 - 739	PV
3		776 - 799	PV
4		856 - 895	PV
5		906 - 951	PV
6		986 - 1021	PV
7	1072 - 1075	1051 - 1088	PV
8		1129 - 1138	PV
9		1157 - 1180	PV
10	1246 - 1249	1223 - 1274	PV
11		1423 - 1428	PV
12		1689 - 1728	PV
13	1826 - 1831	1817 - 1850	PV
14	1980 - 1987		FN
15		2011 - 2012	FP
16		2080 - 2089	FP
17		2098 - 2123	PV
18		2162 - 2199	PV
19		2260 - 2261	PV
20		2278 - 2319	PV
21	2387 - 2391	2368 - 2399	PV
22	2474 - 2479	2456 - 2483	PV
23	2634 - 2637	2610 - 2643	PV

Análogo ao estudo de movimento de clarinetes, aqui também se pode observar grande diferença de resultados entre rotulação e algoritmo. O que poderia aparentemente ser um mau resultado do algoritmo, acaba sendo ao contrário, uma vez que a verificação dos resultados do método, mostrado na Tabela 10, geraram apenas três falhas. Esse resultado serve para ilustrar a diferença entre identificar movimentos via observações e tratá-los matematicamente, via o método proposto.

Ainda de forma similar ao estudo dos movimentos de clarinete e diferente da detecção de movimentos dos joelhos, nota-se que o intervalo sugerido pelo algoritmo novamente não indica a duração do movimento. Como está limiarizado, o método só detecta o intervalo de quadros em que a inclinação das costas foi superior ao limiar, indicando a presença de movimento. Inícios e fins de movimentos,

onde a área calculada é inferior ao fator de limiarização são descartados, reduzindo a duração do gesto, no entanto, gestos mais demorados sugerem maior tempo com áreas superiores a L , gerando um movimento mais demorado no gráfico, conforme ilustra a Figura 41.

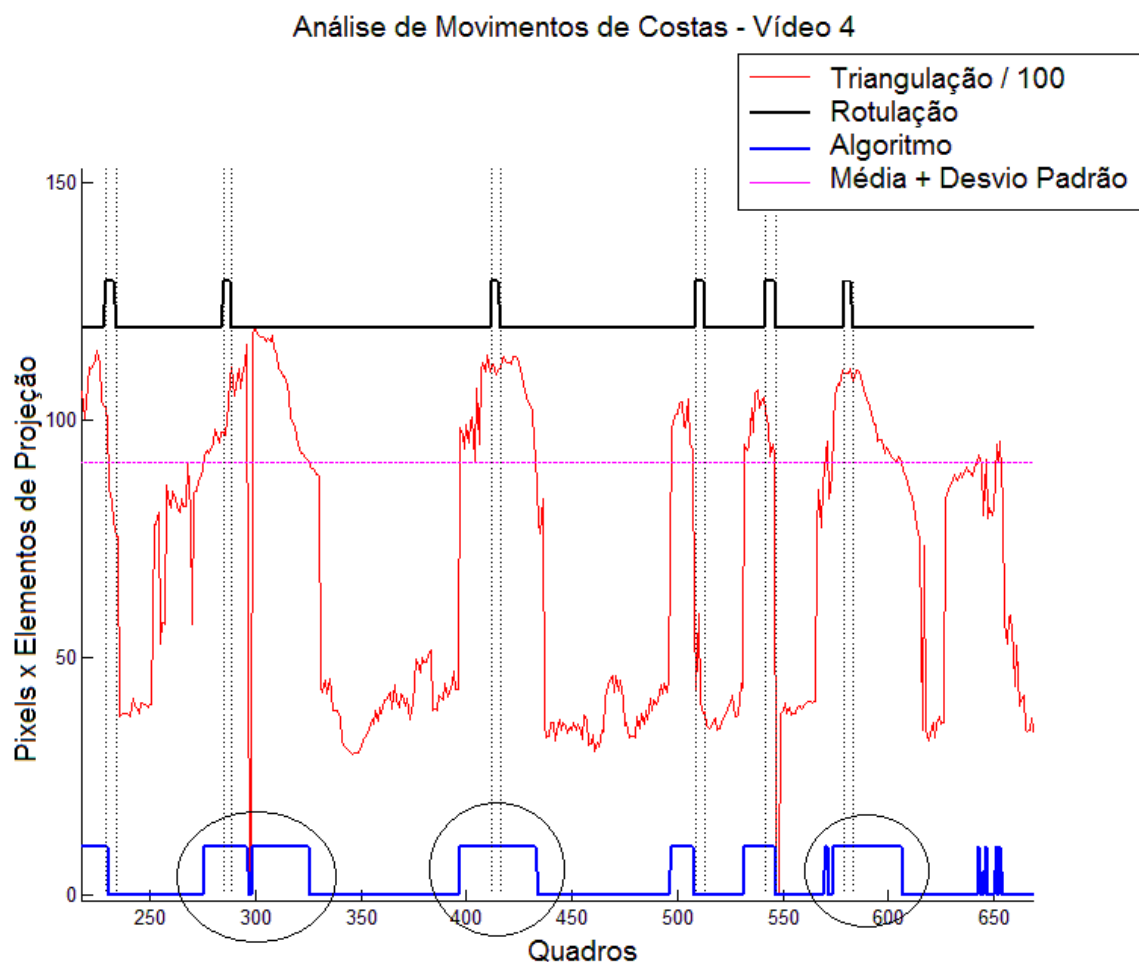


Figura 41: Pulsos alongados no vídeo 4.

Outro padrão ocorrido no estudo e também gerado pela limiarização foi a indução a atrasos de detecção. Conforme ilustra a Figura 42, isso ocorre novamente pelo fato do algoritmo só indicar a presença de gesto quando a área calculada é superior ao limite L . Logo, os inícios demoram alguns quadros para serem observados, causando esta espécie de atraso.

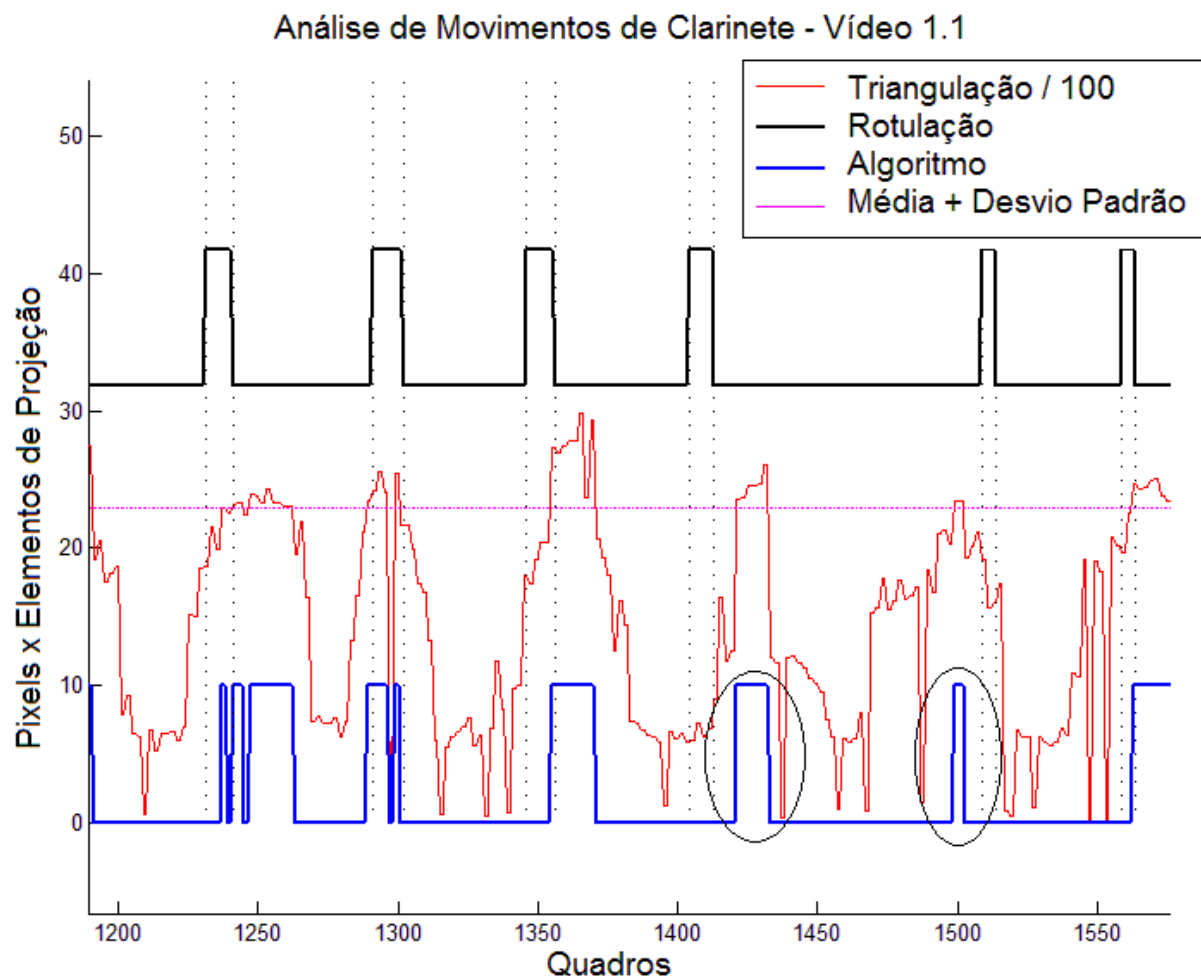


Figura 42: Atraso de detecção no vídeo 1.1.

Já a Figura 43 exemplifica a também ocorrência dos “trens de movimentos” para as costas. Isso acontece quando o movimento encontra-se muito próximo à faixa do limiar, de forma também similar aos movimentos de clarinete. Uma sutil variação na área entre os quadros e esta sendo próximo a L pode descaracterizar o movimento, mesmo ele supostamente estando durante a execução de um gesto. Outro fator que pode causar tal efeito é a presença de quadros com problemas, onde não há cálculo de *bounding box* ou mesmo da triangulação.

Para simplificar a análise, de forma análoga à feita com os clarinetes, regiões onde ocorreram os trens de movimentos foram avaliadas e corrigidas para apenas um único elemento.

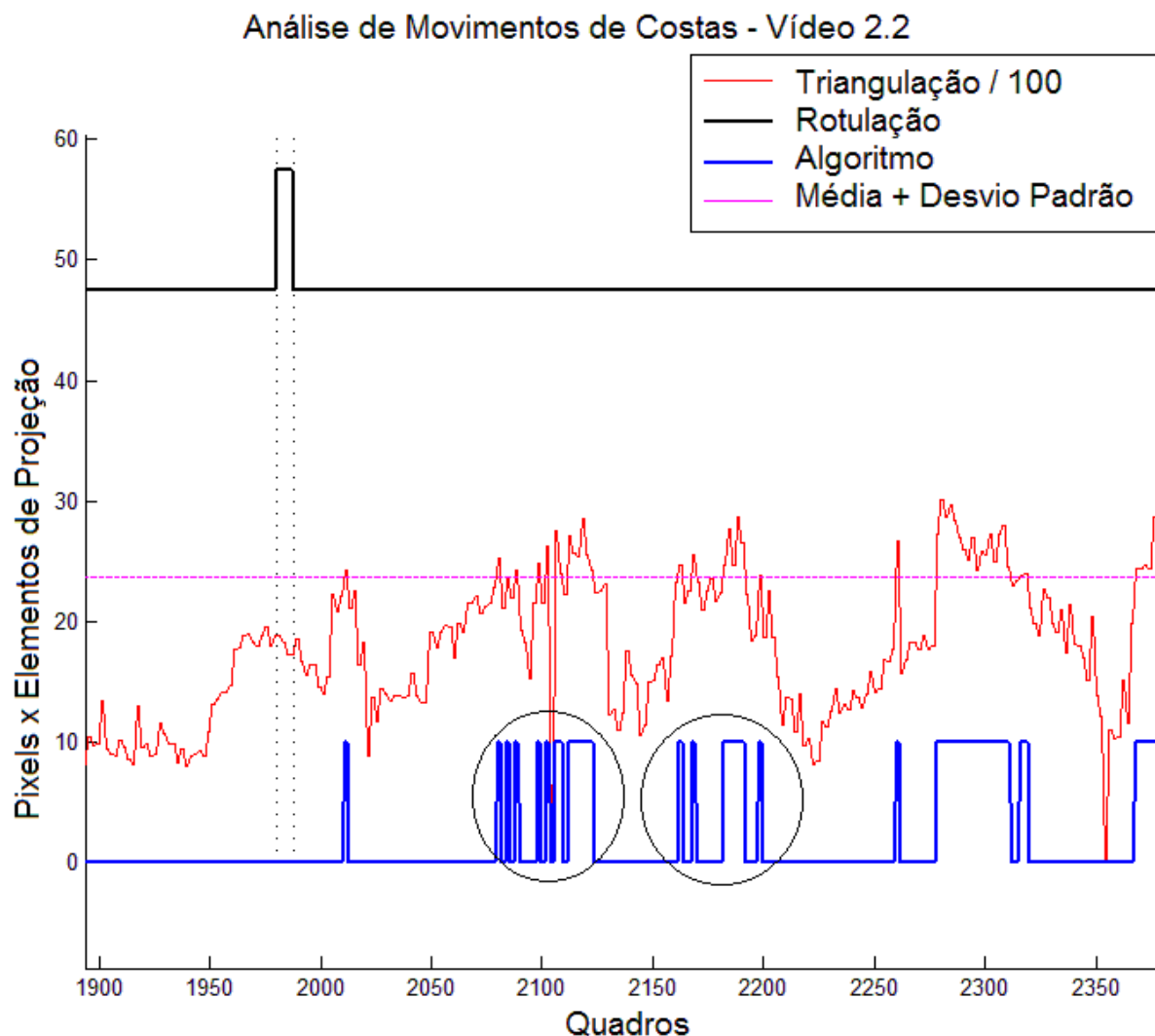


Figura 43: Trem de movimentos no vídeo 2.2.

Reverendo a Tabela 10, percebe-se que dos vinte e três movimentos em questão, apenas três são falsos, sendo dois falsos positivos e um único falso negativo. Os três erros de identificação ocorreram praticamente no mesmo instante, nos trechos entre 66 e 69 segundos do vídeo. A não detecção ocorreu pelo simples fato da inclinação não ter sido o suficiente para alcançar o limiar. Além disso, o primeiro falso positivo ocorreu logo em seguida, na continuação do movimento, no instante em que o clarinetista 2 retira o braço esquerdo do instrumento musical. Como há a falha de formação do ombro no pré-processamento, isto gerou um pico errôneo, mas que, de certa forma, compensou o não alcance do limiar no instante exatamente anterior. Pode-se, então dizer que houve mais um atraso de detecção que propriamente uma falha, porém, este atraso só ocorreu pela má formação do vídeo. O segundo falso positivo também vem na carona do primeiro: a retirada do

braço esquerdo provoca mudança de posicionamento do ombro, causando nova falha de detecção.

No Apêndice, encontram-se referenciados da Tabela 25 a Tabela 31 e da Figura 77 a Figura 84 todos os resultados de análise da base de dados. A Tabela 11 a seguir resume a análise de movimentos auxiliares de ondulações das costas.

Tabela 11: Resultado final da análise de movimentos de ondulação das costas.

ALGORITMO						
ROTULAÇÃO	POSITIVOS VERDADEIROS	FALSOS POSITIVOS	TOTAL DE DETECÇÕES	FALSOS NEGATIVOS	TOTAL DE GESTOS	TAXA DE ACERTO
130	154	12	166	20	186	82.8%

Dentre os três movimentos propostos neste trabalho, o movimento de costas foi o que em números absolutos mais apresentou falhas por não detecção, todavia, em percentual de movimentos, foi de 12%, um valor intermediário.

Outro número interessante é a baixa quantia de falsos positivos, apenas 12 de um total de 166 detecções. Esses baixos valores culminaram com uma alta taxa de acerto, superior a 80%, em se tratando de quase 200 movimentos envolvidos, a maior dentre os três métodos de detecção.

Aproximadamente 85% das falhas de não detecção ocorreram nos vídeos do clarinetista 1. Esses erros existiram pelo mesmo motivo explicado na análise de joelhos: a angulação do músico perante a câmera. Os erros ocorreram pelo fato do músico ter mudado seu eixo de atuação. Esse fator foi mais grave para o movimento de costas, pois a angulação diminui excessivamente o pico de máximo do contorno. Além do mais, a projeção nas duas dimensões da câmera não aumenta significativamente a quantia de elementos do perfil, mantendo a área da triangulação inferior ao limiar. Nos joelhos, a angulação gerou falsos positivos, já para as costas, o efeito foi o contrário.

Os erros de falso-positivo ocorreram devido aos músicos não retornarem a um ponto estável após o movimento. O clarinetista executa o movimento, mas retorna a um ponto muito próximo do referencial, gerando a imperfeição na medida, variação na região de atuação do limiar. Outros erros ocorreram ainda por problemas de pré-processamento (clarinetista 2) ou confusão com outro movimento: gestos de cabeça para o clarinetista 3 e dobra de joelhos para o clarinetista 4.

5 CONCLUSÃO

Avaliando a metodologia proposta e os resultados obtidos, conclui-se que é possível a união de diversas técnicas clássicas de visão computacional com o intuito de detecção de movimentos auxiliares específicos produzidos por clarinetistas.

A alta taxa de acerto para o movimento pendular de clarinete – 76% – foi de certa forma, um resultado já esperado, pelo fato da metodologia ser relativamente intuitiva. Esta taxa seria ainda maior, de cerca de 90%, se não houvesse problemas com os vídeos do clarinetista 2 – 80% das falhas. As poucas falhas de detecção ocorreram basicamente pela escolha do limiar da curva de diferenças, ou seja, o ajuste, a sensibilidade do algoritmo reflete diretamente na capacidade de identificação. Uma possível forma de se atenuar esta influência (trabalho futuro) seria o cálculo de máximos locais na curva de diferença absoluta.

Os quase 70% de taxa de acerto na detecção de gestos auxiliares de dobra de joelhos foram de certa forma expressivos. Este movimento é muito difícil de ser detectado, uma vez que a sua execução gera influência em outras partes do corpo, além do fato de outros movimentos, como deslocamento de peso, inclinação do corpo e passos também influenciarem diretamente na sua identificação. Outro fator complicador para a análise é o fato de o gesto ser produzido por duas partes distintas do corpo. Logo, a forma como o vídeo é gerado também implica na obtenção de resultados. Mesmo assim, a proposta de análise da projeção de perfil do local mostrou-se acertada.

A metodologia para avaliação dos gestos auxiliares de ondulação das costas foi a que atingiu a maior porcentagem de acerto – superior a 80% –, porém foi também a que apresentou maior índice de não detecção. Este tipo de erro não é desejável pelo fato da proposta deste trabalho ser a substituição das repetidas observações dos vídeos, pela simples varredura do *software*. Uma “não detecção humana” é passível de correção através de repetidas observações, já o algoritmo é determinístico, sempre apresenta o mesmo resultado, a não ser que receba realimentação, ou ajuste.

Mesmo assim, um atenuador para este resultado é o fato dos erros de não detecção terem ocorrido substancialmente por problemas na aquisição de dados.

85% das falhas ocorreram pela mudança do posicionamento do músico em relação à câmera, enquanto que apenas 15% (três erros) foram por confusão com outro tipo de movimento.

Em Wanderley, et. al. 2005 há estudo de movimentos baseado em um dos vídeos do clarinetista 1, assim como o feito nesta pesquisa. Neste trabalho foram rotulados 35 movimentos de clarinete, 34 movimentos de ondulações nas costas (na referência foram denominados movimentos de cabeça) e nenhum movimento de dobra de joelhos. A rotulação prévia feita nesta pesquisa observou 34 movimentos de clarinete, 32 movimentos de costas e nenhum movimento de joelhos, ao passo que a metodologia proposta identificou 38 gestos de clarinete, 22 de costas e nenhum de joelhos.

Nota-se uma forte relação entre os resultados e esta comparação ilustra os bons números obtidos pelo método proposto. O algoritmo sugeriu poucos movimentos de clarinete a mais, que numa segunda observação foram confirmados, no entanto não detectou cerca de 10 movimentos de costas por problemas que já foram justificados. Já os resultados para os gestos de dobra de joelhos foram idênticos.

Outro detalhe comum nas análises foi a formação de trem de movimentos. Tanto para as costas, como para o clarinete, este efeito ocorreu com frequência, mas pode também ser neutralizado. A própria detecção de máximos locais na curva de diferenças poderia atenuar este efeito. Já para as costas, o cálculo da taxa de variação da área, agregado à limiarização, poderia amenizar esta influência. Estas são sugestões de trabalhos futuros.

O que também se permite concluir é que o pré-processamento dos vídeos é de suma importância. Todos os resultados obtidos estão intimamente ligados à qualidade imposta pela etapa inicial. Logo, a melhoria destes resultados passa obrigatoriamente pelo aprimoramento desta primeira etapa. Além disso, fazer a aquisição de vídeos já conhecendo as limitações da etapa de pré-processamento também colaboraria para melhores resultados. Por exemplo, obtendo novos vídeos para o clarinetista 2, com um traje similar ao dos outros instrumentistas, certamente geraria melhores resultados quanto às detecções de movimentos.

A limiarização matemática como indicador de tomadas de decisão possibilitou altas taxas de acerto, porém não da forma ideal. Quando descartadas questões de

obtenção dos vídeos e/ou de pré-processamento, grande parte dos erros de detecção foi devido a esta técnica. A forte necessidade de ajuste de parâmetros vídeo a vídeo pode ser considerada como a limitação desta pesquisa. A melhoria destes resultados também serve como sugestão de trabalhos futuros.

Enfim, mesmo com estes atenuantes, o objetivo proposto, de avaliar os vídeos de clarinetistas disponíveis e a partir somente destes identificar movimentos pendulares de clarinete, dobra de joelhos e ondulações nas costas foi alcançado. As técnicas estudadas apresentaram resultado satisfatório, ao passo que suas limitações são também conhecidas.

Pode-se questionar se todo o trabalho necessário no desenvolvimento da metodologia é vantajoso perante o esforço envolvido na observação dos movimentos – principal objetivo deste estudo. Mesmo sendo necessário o estudo de diferentes técnicas, criação de algoritmos e ajustes de diversos parâmetros, os resultados comprovam a validade da ideia aqui fundamentada.

Para os três movimentos propostos, foi possível atingir taxas de acerto superiores a 70% e grande parte dos erros de detecção foi provocada pela forma como os vídeos foram originados. Além disso, o algoritmo sempre sugeriu uma quantidade de gestos acima dos rotulados previamente.

É natural que uma pesquisa em caráter inicial demande muito esforço para atingir seu resultado esperado. No entanto, as limitações conhecidas, assim como as propostas sugeridas permitem que trabalhos futuros aprimorem os resultados finais, ratificando a validade do estudo.

Como proposta futura, podem-se aplicar estas melhorias sugeridas e, então, reunir todos os métodos estudados em um único *software* identificador de movimentos. Este programa indicaria intervalos de tempo dos vídeos que contêm gestos, facilitando a filtragem, não mais necessitando as inúmeras observações dos vídeos, aprimorando o processo para futuras análises mais detalhadas.

REFERÊNCIAS

AGGARWAL, J. K.; CAI, Q. Human Motion Analysis: A Review. In: **Computer Vision and Image Understanding**. Wadern: Schloss Dagstuhl, 1999. p. 428-440.

BANDERA, J. P.; MARFIL, R.; BANDERA, A.; RODRÍGUEZ, J. A.; MOLINA-TANCO, L.; SANDOVAL, F. Fast Gesture Recognition Based on a Two-level Representation. **Pattern Recognition Letters**. Elsevier, n. 30, p. 1181-1189, 2009.

BOBICK, A. F.; DAVIS, J. W. The Recognition of Human Movement Using Temporal Templates. **IEEE Transactions on Pattern Analysis and Machine Intelligence**. [S.l.], v. 23, n. 3, p. 257-267, 2001.

BRADSKI, G.; KAEHLER, A. **Learning OpenCV: Computer Vision with the OpenCV Library**. Sebastopol: O'Reilly, 2008. 556 p.

CADOZ, C. Instrumental Gesture and Music Composition. In: INTERNATIONAL COMPUTER MUSIC CONFERENCE, 1988, Colônia, Alemanha. **Proceedings of the International Computer Music Conference**. São Francisco: ICMA, 1988.

CADOZ, C.; WANDERLEY, M. M. Gesture: music. In: WANDERLEY, M. M.; BATTIER, M. (Eds.) **Trends in Gestural Control of Music**. Paris: IRCAM/Centre Pompidou, 2000.

DAHL, S.; FRIBERG, A. Expressiveness of Musician's Body Movements in Performances on Marimba. In: CAMURRI, A.; VOLPE, G. (Eds.) **Gesture-Based Communication in Human-Computer Interaction**. Berlim: Springer-Verlag, 2004. p. 479-486.

DAVIDSON, J. W. Visual Perception of Performance Manner in the Movements of Solo Musicians. **Psychology of Music**. Londres, n. 21, p. 103-113, 1993.

DELALANDE, F. La Gestique de Gould: Éléments pour une Sémiologie du Geste Musical. In: GERTIN, G (Ed.). **Glenn Gould Pluriel**. Quebec: Louise Courteau Editrice, 1988. p. 85-111.

GAVRILA, D. M. The Visual Analysis of Human Movement: A Survey. In: **Computer Vision and Image Understanding**. Wadern: Schloss Dagstuhl, 1999. p. 82-98.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing**. 3.ed. Upper Saddle River: Prentice Hall, 2007. 976p.

GUNES, H.; PICCARDI, M. Bi-modal Emotion Recognition from Expressive Face and Body Gestures. **Journal of Network and Computer Applications**. Londres, n. 30, p. 1334-1345, 2007.

HUMMELS, C.; SMETS, G.; OVERBEEKE, K. An Intuitive Two-Handed Gestural Interface for Computer Supported Product Design. In: WACHSMUTH, I.; FRÖLICH, M. **Gesture and Sign Language in Human-Computer Interaction**. Berlin: Springer-Verlag, 1998. p. 197-208.

JENSENIUS, A. R.; WANDERLEY, M. M; GODOY, R. I; LEMAN, M. Musical Gesture: Concepts and Methods in Research. In: GODOY, R. I.; LEMAN, M. (Eds.). **Musical Gestures: Sound, Movement and Meaning**. Nova Iorque: Routledge, 2010. p.12-35.

KOERICH, A. L. **Uma Proposta para o Processamento Automático de Cheques Bancários: Extração e Armazenamento de Informações**. 1997. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Estadual de Campinas, UNICAMP, Campinas, 1997.

KURTENBACH, G; HULTEEN, E. Gestures in Human-Computer Communication. In LAUREL, B (Ed.). **The Art of Human-Computer Interface Design**. Ed Addison Wesley. 1990.

LI, H.; GREENSPAN, M. Model-Based Segmentation and Recognition of Dynamic Gestures in Continuous Video Streams. **Pattern Recognition**. Elsevier, n. 44, p.1614-1628, 2011.

LUCAS, B. D; KANADE, T. An Iterative Image Registration Technique with an Application to Stereo Vision. In: IMAGE UNDERSTANDING WORKSHOP, 1981, Washington, Estados Unidos. **Proceedings of Imaging Understanding Workshop**. [S.I.], 1981. p. 121–130.

MAZZARINO, B.; PEINADO, M.; BOULIC, R.; VOLPE, G.; WANDERLEY, M. M. Improving the Believability of Virtual Characters Using Qualitative Gesture Analysis. In: DIAS, M. S.; GIBET, S.; WANDERLEY, M. M.; BASTOS, R. (Eds.). **Gesture-Based Human-Computer Interaction and Simulation**. Berlim: Springer-Verlag, 2009. p. 48-56.

MITRA, S.; ACHARYA, T. Gesture Recognition: A Survey. **IEEE Transactions on Systems, Man and Cybernetics**. [S.l.], v. 37, n. 3, p. 311-324, 2007.

ROSENBLUM, M.; YACOOB, Y.; DAVIS, L. S. Human Expression Recognition from Motion Using a Radial Basis Function Network Architecture. **IEEE Transactions on Neural Networks**. [S.l.], v. 7, n. 5, p. 1121-1138, 1996.

SHAH, M. **Fundamentals of Computer Vision**. Disponível em: <<http://www.math.ucf.edu/csums/BOOK.PDF>>. Acesso em: 19 out. 2010.

STEINBRUCH, A.; WINTERLE, P. **Geometria Analítica**. 2.ed. São Paulo: McGraw-Hill, 1987. 291p.

SZELISKI, R. **Computer Vision: Algorithms and Applications**. Disponível em: <<http://szeliski.org/Book/>>. Acesso em: 18 out. 2010.

TEIXEIRA, E. C. F. **Análise da Expressividade Musical com Base em Medidas Acústicas e do Gesto Físico**. 2010. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Federal de Minas Gerais, UFMG, Belo Horizonte, 2010.

VERFAILLE, V.; QUEK, O.; WANDERLEY, M. M. Sonification of Musicians' Ancillary Gestures. In: INTERNATIONAL CONFERENCE ON AUDITORY DISPLAY, 2006, Londres, Inglaterra. **Proceedings of the 12th International Conference on Auditory Display**. [S.l.], 2006.

WANDERLEY, M. M. Non-obvious Performer Gestures in Instrumental Music. In: BRAFFORT, A.; GHERBI, R.; GIBET, S.; RICHARDSON, J.; TEIL, D (Eds.). **Gesture-based Communication in Human-Computer Interaction**. Berlim: Springer-Verlag, 1999. p. 37-48.

WANDERLEY, M. M. Quantitative Analysis of Non-Obvious Performer Gestures. In: WACHSMUTH, I.; SOWA, T (Eds.). **Gesture and Sign Language in Human-Computer Interaction**. Berlim: Springer-Verlag, 2002. p. 241-253.

WANDERLEY, M. M.; VINES, B. W.; MIDDLETON, N.; McKAY, C.; HATCH, W. The Musical Significance of Clarinetists' Ancillary Gestures: An Exploration of the Field. **Journal of New Music Research**. Londres, v. 34, n. 1, p. 97-113, 2005.

WILSON, A. D.; BOBICK, A. F.; CASSELL, J. Recovering the Temporal Structure of Natural Gesture. In: AUTOMATIC FACE AND GESTURE RECOGNITION, Killington, Estados Unidos, 1996. **Proceedings of the Second International Conference on Automatic Face and Gesture Recognition**. [S.I.], 1996. p.XVI-XXVI.

ZRAMDINI, A.; INGOLD, R. Optical Font Recognition from Projection Profiles. **Electronic Publishing**. [S.I.], v. 6, n. 3, p. 249-260, 1993.

APÊNDICE

Tabela 12: Resultado da análise do movimento de clarinete para o vídeo 1.1.

GESTO	Intervalo de Quadros		RESULTADO
	ROTULAÇÃO	ALGORITMO	
1		79 - 107	PV
2	134 - 158	138 - 165	PV
3	189 - 225	200 - 209	PV
4	261 - 278	260 - 275	PV
5	311 - 331	304 - 325	PV
6	363 - 388	362 - 379	PV
7		467 - 471	PV
8	526 - 549	548 - 549	PV
9	579 - 605	583 - 605	PV
10	647 - 669	640 - 663	PV
11	708 - 732	712 - 737	PV
12	790 - 805	774 - 793	PV
13	839 - 853	822 - 845	PV
14		872 - 874	PV
15	936 - 958	934 - 967	PV
16	1088 - 1112	1090 - 1109	PV
17	1139 - 1169	1151 - 1161	PV
18	1210 - 1227	1208 - 1227	PV
19	1273 - 1291	1269 - 1291	PV
20	1311 - 1334	1309 - 1330	PV
21	1378 - 1410	1384 - 1408	PV
22	1451 - 1480	1453 - 1498	PV
23		1507 - 1566	PV
24	1592 - 1618	1593 - 1616	PV
25	1641 - 1667	1641 - 1662	PV
26	1696 - 1723	1701 - 1728	PV
27	1763 - 1784	1757 - 1778	PV
28	1806 - 1836	1809 - 1832	PV
29	1864 - 1891	1879 - 1880	PV
30	1924 - 1950	1927 - 1942	PV
31	1982 - 2010		FN
32	2036 - 2063	2029 - 2063	PV
33	2105 - 2129	2102 - 2121	PV
34	2163 - 2196	2168 - 2183	PV

Tabela 13: Resultado da análise do movimento de clarinete para o vídeo 1.2.

GESTO	Intervalo de Quadros		RESULTADO
	ROTULAÇÃO	ALGORITMO	
1		201	PV
2	296 - 325	298 - 332	PV
3		335 - 344	PV
4	375 - 403	380 - 396	PV
5		458 - 463	PV
6	521 - 545	514 - 549	PV
7	608 - 625	609 - 614	PV
8	664 - 686	671 - 672	PV
9		722 - 734	PV
10	790 - 819	793 - 820	PV
11		866 - 870	PV
12		952 - 993	PV
13	1052 - 1080	1048 - 1072	PV
14	1107 - 1135	1111 - 1136	PV
15		1177 - 1199	PV
16	1240 - 1265	1234 - 1277	PV
17		1279 - 1299	FP
18	1390 - 1414	1391 - 1473	PV
19		1476 - 1484	PV
20	1565 - 1584	1548 - 1590	PV
21	1642 - 1669	1619 - 1689	PV
22	1722 - 1746	1720 - 1748	PV
23		1798 - 1810	PV
24	1856 - 1883	1855 - 1873	PV
25		1932 - 1996	PV
26		2048 - 2082	PV
27		2085 - 2108	FP
28		2118 - 2144	FP

Tabela 14: Resultado da análise do movimento de clarinete para o vídeo 1.3.

GESTO	Intervalo de Quadros		RESULTADO
	ROTULAÇÃO	ALGORITMO	
1		583 - 584	FP
2		644 - 651	PV
3	685 - 717		FN
4	747 - 777	758 - 775	PV
5	805 - 822	802 - 821	PV
6	846 - 862	833 - 857	PV
7	888 - 912	887 - 913	PV
8		951 - 977	PV
9	1011 - 1036	1008 - 1039	PV
10	1071 - 1089	1062 - 1080	PV
11	1116 - 1134	1099 - 1132	PV
12	1166 - 1190	1161 - 1192	PV
13	1221 - 1240	1215 - 1238	PV
14	1264 - 1288	1261 - 1284	PV
15	1309 - 1335	1305 - 1338	PV
16	1385 - 1417	1373 - 1446	PV
17	1477 - 1501	1477 - 1500	PV
18	1523 - 1553	1525 - 1554	PV
19	1579 - 1612	1583 - 1615	PV
20	1651 - 1671	1649 - 1670	PV
21	1689 - 1713	1689 - 1718	PV
22	1745 - 1768	1742 - 1774	PV
23	1819 - 1846	1814 - 1844	PV
24	1868 - 1897	1867 - 1911	PV
25		1937 - 1964	FP
26	1991 - 2015	1991 - 2012	PV
27	2032 - 2060	2035 - 2065	PV
28	2094 - 2129	2102 - 2135	PV
29	2163 - 2189	2160 - 2187	PV
30	2216 - 2243	2210 - 2249	PV
31	2278 - 2305	2278 - 2303	PV
32	2340 - 2361	2342 - 2357	PV
33	2439 - 2477	2447 - 2473	PV
34	2515 - 2546	2522 - 2529	PV
35	2566 - 2601	2575 - 2609	PV

Tabela 15: Resultado da análise do movimento de clarinete para o vídeo 2.1.

GESTO	Intervalo de Quadros		RESULTADO
	ROTULAÇÃO	ALGORITMO	
1		273 - 274	FP
2		300	FP
3	351 - 362	353 - 354	PV
4	522 - 531	515 - 537	PV
5		550 - 552	FP
6		597	FP
7		609 - 612	PV
8	831 - 844	807 - 811	PV
9		843 - 849	FP
10		935 - 945	FP
11		955 - 964	FP
12		969 - 978	FP
13		983 - 998	FP
14		1001 - 1024	PV
15		1027 - 1035	FP
16		1039 - 1076	FP
17		1079 - 1100	FP
18		1102 - 1113	FP
19		1143 - 1156	PV
20		1428 - 1439	PV
21		1620 - 1632	PV
22		1729 - 1759	FP
23		1798 - 1818	FP
24		1851 - 1863	FP
25		1872 - 1888	PV
26		1900 - 1907	FP
27		1934 - 1957	FP
28		1974 - 2003	FP
29	2015 - 2021	2048 - 2062	PV
30		2064 - 2095	FP
31		2097 - 2114	FP
32		2127 - 2128	FP
33		2174	FP
34		2204 - 2219	PV
35		2228 - 2241	FP
36		2299 - 2312	FP

Tabela 16: Resultado da análise do movimento de clarinete para o vídeo 2.2.

GESTO	Intervalo de Quadros		RESULTADO
	ROTULAÇÃO	ALGORITMO	
1		418 - 445	FP
2		494 - 516	FP
3		518 - 525	PV
4	558 - 568	570 - 591	PV
5		623 - 624	FP
6		676 - 694	PV
7		700 - 712	PV
8	761 - 771	714 - 744	PV
9	812 - 831	779 - 801	PV
10		854 - 855	FP
11		903 - 917	PV
12	921 - 935	922 - 957	PV
13		1005 - 1021	PV
14		1056 - 1079	PV
15		1107	PV
16	1192 - 1212	1194 - 1214	PV
17		1235 - 1241	FP
18	1306 - 1329	1305 - 1326	PV
19	1353 - 1378		FN
20		2109 - 2112	FP
21		2299 - 2306	FP
22	2401 - 2410	2370 - 2393	PV
23	2461 - 2467	2464 - 2479	PV
24		2624 - 2630	PV
25		2713 - 2759	FP

Tabela 17: Resultado da análise do movimento de clarinete para o vídeo 2.3.

GESTO	Intervalo de Quadros		RESULTADO
	ROTULAÇÃO	ALGORITMO	
1		74 - 232	FP
2	228 - 245	231 - 232	PV
3		254 - 298	FP
4	336 - 347	341 - 353	PV
5	371 - 384		FN
6	445 - 459	420 - 437	PV
7	496 - 517		FN
8		551 - 556	FP
9		563 - 621	PV
10		637 - 652	FP
11	728 - 748	662 - 721	PV
12		756 - 801	PV
13	845 - 856	828 - 837	PV
14	907 - 926	863 - 889	PV
15		944 - 946	PV
16	1021 - 1039		FN
17	1072 - 1095		FN
18		1114 - 1147	FP
19		1154 - 1172	FP
20		1209 - 1225	PV
21		1228 - 1271	FP
22		1274 - 1301	FP
23		1512 - 1543	FP
24		1546 - 1555	FP
25		1608 - 1714	FP
26		1799 - 1826	FP
27		1874 - 1885	PV
28	2150 - 2159		FN
29	2321 - 2336	2301 - 2317	PV

Tabela 18: Resultado da análise do movimento de clarinete para o vídeo 3.

GESTO	Intervalo de Quadros		RESULTADO
	ROTULAÇÃO	ALGORITMO	
1	347 - 376	366 - 377	PV
2	559 - 574	563 - 574	PV
3	666 - 681	671 - 673	PV
4	807 - 823	798 - 820	PV
5		1049	FP
6	1206 - 1221		FN
7		1253 - 1288	PV
8	1299 - 1322	1324 - 1345	PV
9		1382 - 1425	PV
10		1548 - 1549	PV
11	1729 - 1745	1721 - 1737	PV
12		1801 - 1857	FP
13	1879 - 1917	1862 - 1924	PV
14	1982 - 2030	1978 - 2021	PV

Tabela 19: Resultado da análise do movimento de clarinete para o vídeo 4.

GESTO	Intervalo de Quadros		RESULTADO
	ROTULAÇÃO	ALGORITMO	
1	167 - 178	168 - 181	PV
2	232 - 241	236	PV
3	342 - 353	343 - 355	PV
4	436 - 445	437 - 482	PV
5	514 - 522	513 - 518	PV
6	581 - 595	589 - 599	PV
7		613 - 625	PV
8		666 - 679	PV
9		836 - 839	FP
10		875 - 883	PV
11		957 - 959	FP
12		1009 - 1033	FP
13		1047 - 1072	FP
14		1114 - 1134	PV
15		1301 - 1309	PV
16	1430 - 1438	1430 - 1434	PV
17	1541 - 1550	1537 - 1551	PV
18		1646 - 1686	PV

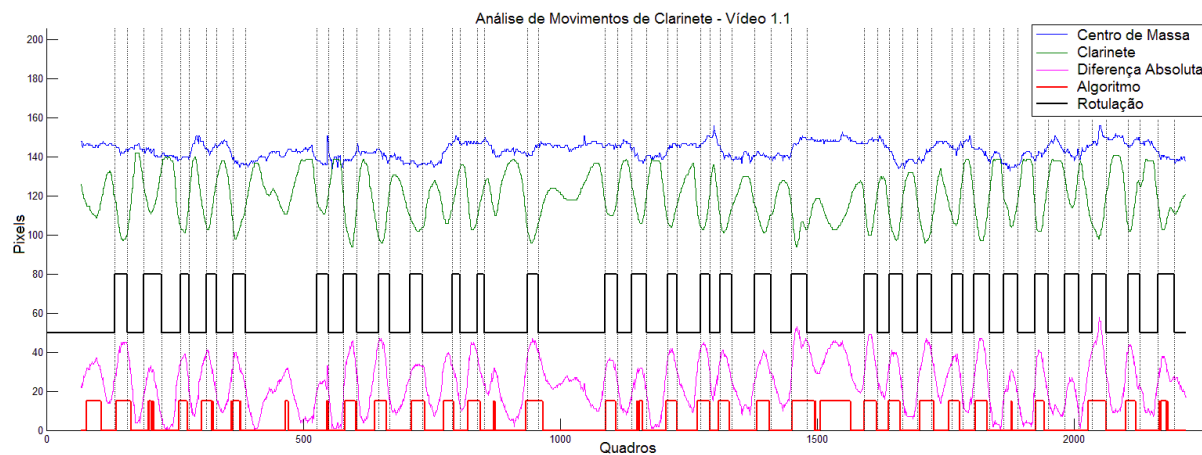


Figura 44: Resultado da análise do movimento de clarinete para o vídeo 1.1.

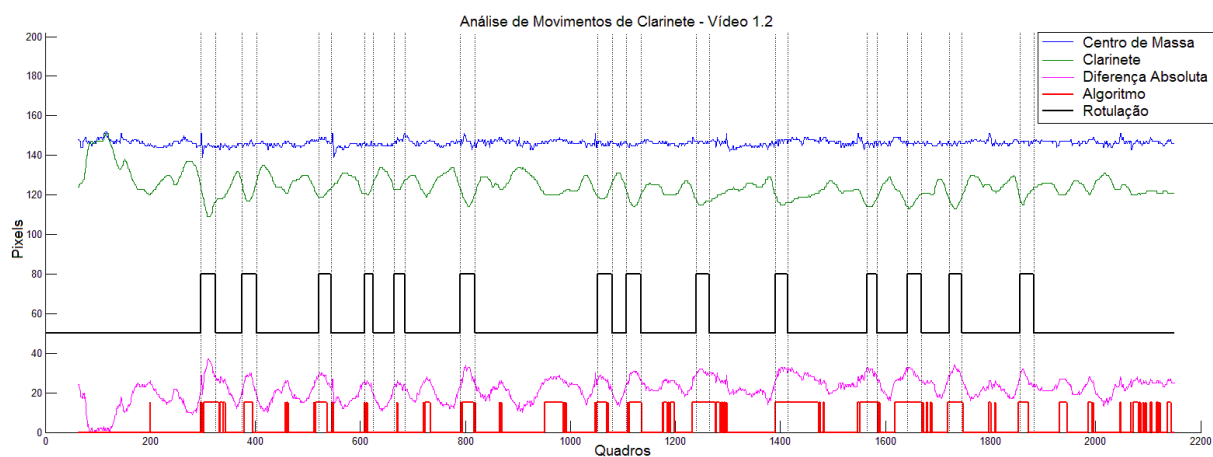


Figura 45: Resultado da análise do movimento de clarinete para o vídeo 1.2.

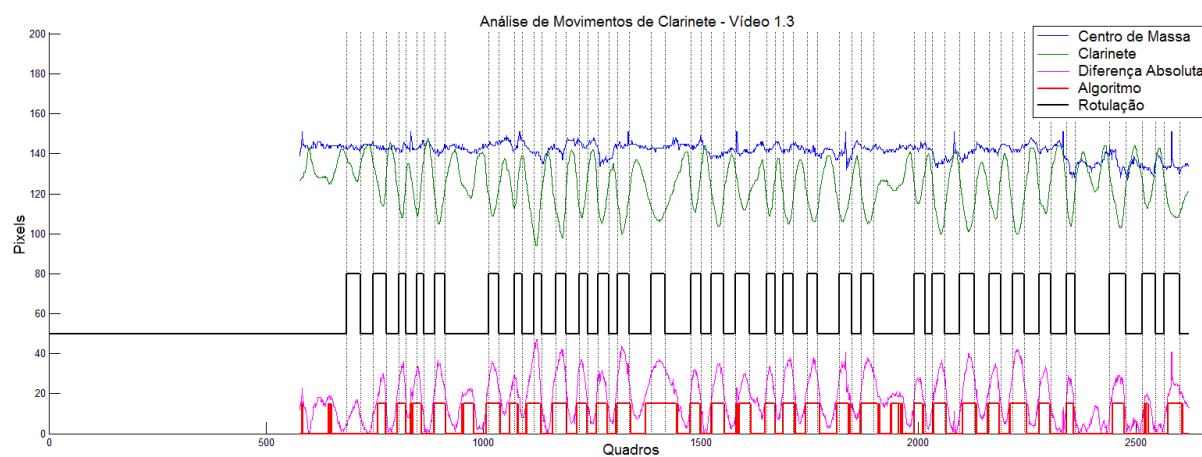


Figura 46: Resultado da análise do movimento de clarinete para o vídeo 1.3.

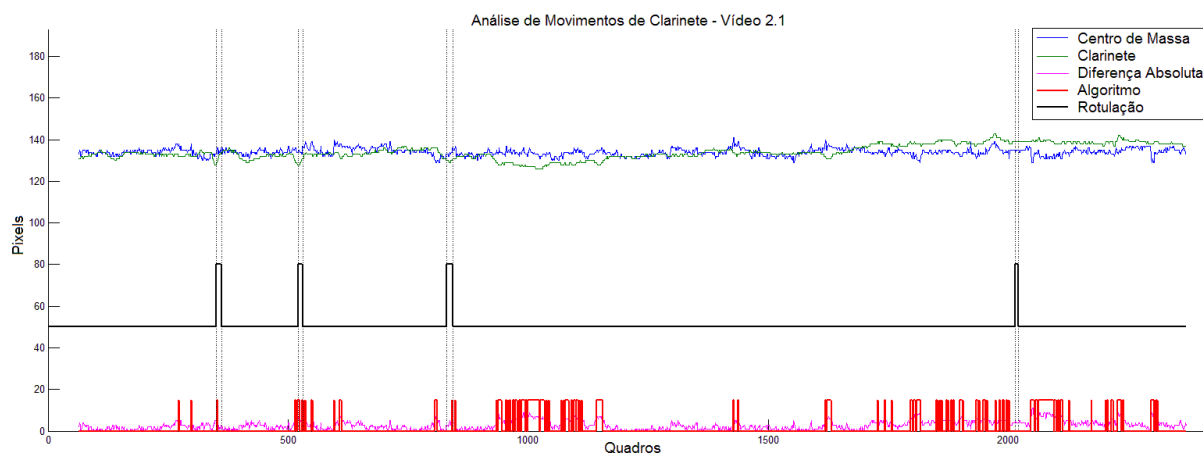


Figura 47: Resultado da análise do movimento de clarinete para o vídeo 2.1.

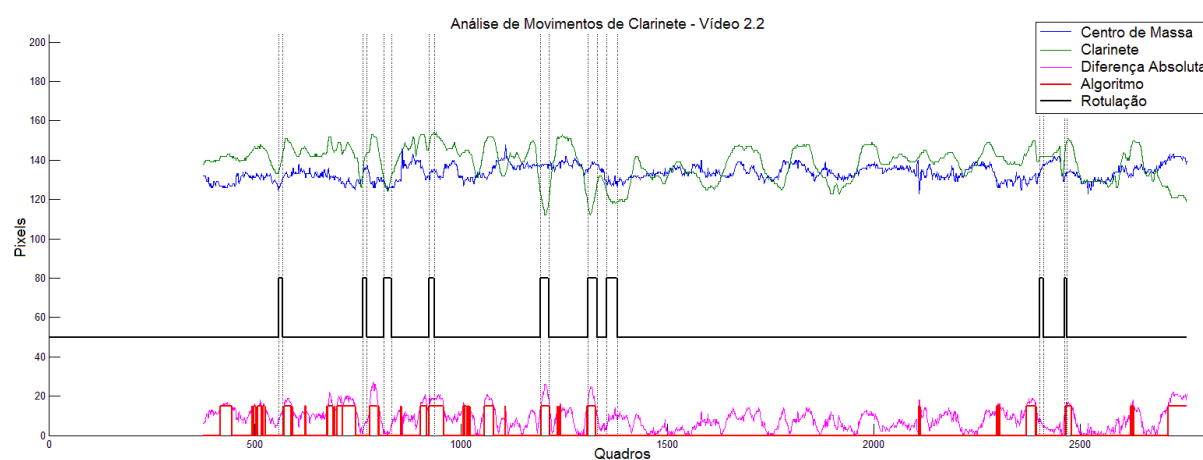


Figura 48: Resultado da análise do movimento de clarinete para o vídeo 2.2.

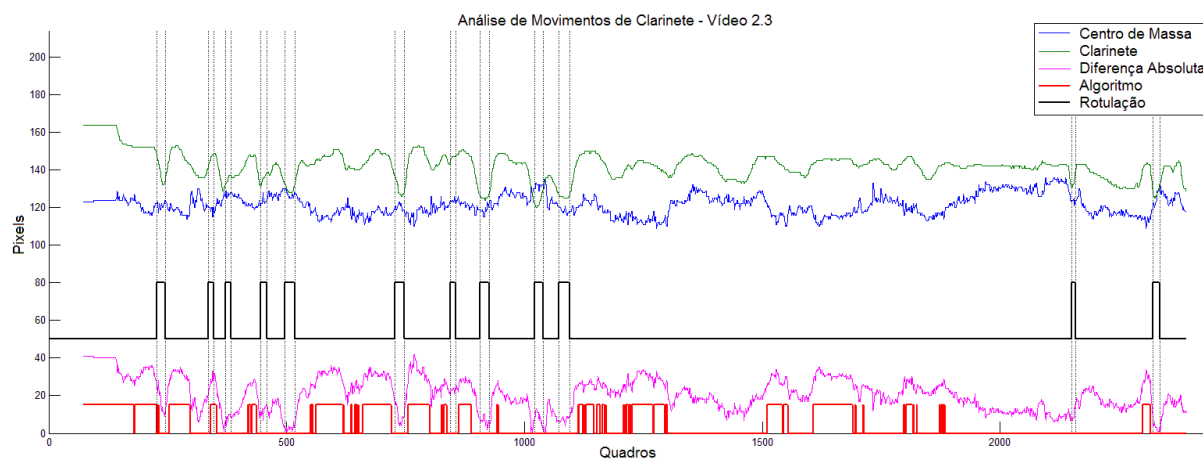


Figura 49: Resultado da análise do movimento de clarinete para o vídeo 2.3.

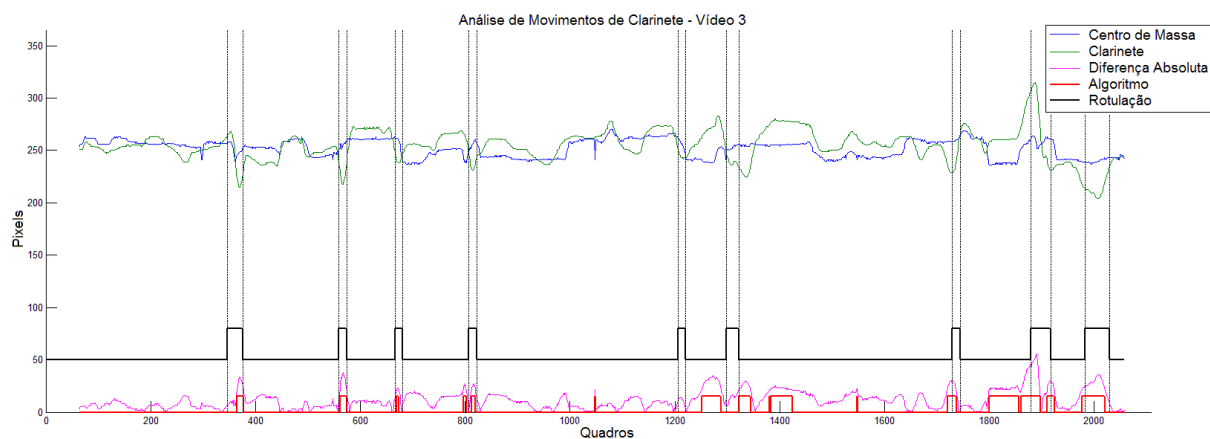


Figura 50: Resultado da análise do movimento de clarinete para o vídeo 3.

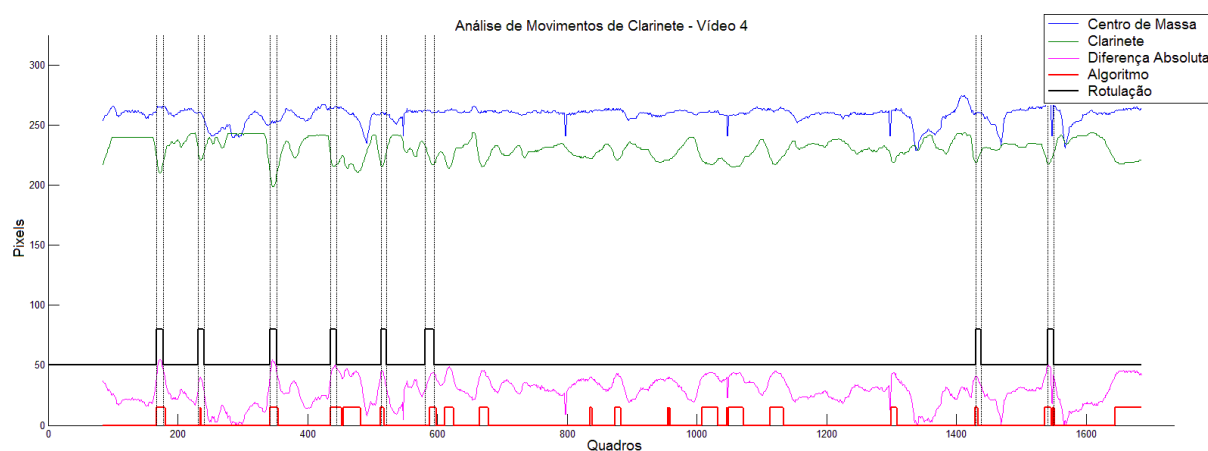


Figura 51: Resultado da análise do movimento de clarinete para o vídeo 4.

Tabela 20: Resultado da análise do movimento de joelhos para o vídeo 1.1.

GESTO	Intervalo de Quadros		RESULTADO
	ROTULAÇÃO	ALGORITMO	
1		701 - 710	FP
2		744 - 755	PV
3		911 - 924	FP
4		928 - 943	FP
5		1014 - 1031	FP
6	1069 - 1077	1053 - 1061	PV
7	2299 - 2304	2289 - 2292	PV

Tabela 21: Resultado da análise do movimento de joelhos para o vídeo 2.1.

GESTO	Intervalo de Quadros		RESULTADO
	ROTULAÇÃO	ALGORITMO	
1	717 - 722	696 - 703	PV
2	1977 - 1984	1965 - 1972	PV

Tabela 22: Resultado da análise do movimento de joelhos para o vídeo 2.2.

GESTO	Intervalo de Quadros		RESULTADO
	ROTULAÇÃO	ALGORITMO	
1	581 - 587	606 - 609	PV
2	678 - 683		FN
3	785 - 788		FN
4	855 - 865		FN
5	989 - 995	971 - 985	PV
6		1046 - 1047	PV
7		1115 - 1122	PV
8	1191 - 1198		FN
9	1239 - 1244		FN
10	1302 - 1308	1284 - 1301	PV
11	1349 - 1355	1325 - 1336	PV
12		1442 - 1562	PV
13	1668 - 1673	1568 - 1679	PV
14	1733 - 1738	1767 - 1804	PV
15	1807 - 1817	1849 - 1941	PV
16	2116 - 2122	2128 - 2139	PV
17		2221 - 2261	PV
18	2319 - 2326	2349 - 2356	PV
19	2407 - 2419	2399 - 2446	PV
20	2452 - 2459	2474 - 2512	PV
21	2593 - 2598	2592 - 2603	PV
22	2656 - 2663	2641 - 2672	PV
23	2678 - 2684	2680 - 2695	PV
24		2700 - 2758	FP

Tabela 23: Resultado da análise do movimento de joelhos para o vídeo 2.3.

GESTO	Intervalo de Quadros		RESULTADO
	ROTULAÇÃO	ALGORITMO	
1		210 - 215	PV
2	355 - 360	361 - 367	PV
3	554 - 560	627 - 634	PV
4	687 - 692	688 - 695	PV
5		713 - 718	FP
6		799 - 815	PV
7		823 - 835	FP
8		845 - 853	PV
9		976 - 1017	PV
10		1167 - 1182	FP
11		1315 - 1332	PV
12		1379 - 1414	FP
13		1443 - 1451	FP
14		1463 - 1473	PV
15		1606 - 1616	FP
16	1717 - 1723		FN
17	1910 - 1917		FN
18	2136 - 2145	2119 - 2143	PV
19		2304 - 2309	PV
20	2352 - 2357	2342 - 2361	PV
21	2384 - 2390		FN

Tabela 24: Resultado da análise do movimento de joelhos para o vídeo 4.

GESTO	Intervalo de Quadros		RESULTADO
	ROTULAÇÃO	ALGORITMO	
1	105 - 118	99 - 103	PV
2	256 - 259		FN
3	310 - 314	310 - 330	PV
4	412 - 417	416 - 430	PV
5		454 - 472	PV
6	494 - 501	498 - 508	PV
7	581 - 585	570 - 574	PV
8	655 - 663	654 - 664	PV
9		842 - 850	PV
10		853 - 864	PV
11	901 - 905		FN
12	1084 - 1088	1073 - 1098	PV
13	1172 - 1177	1159 - 1163	PV
14		1192 - 1224	PV
15	1352 - 1360	1359 - 1362	PV
16	1397 - 1405	1392 - 1424	PV
17		1499 - 1509	FP
18		1527 - 1544	FP
19		1546 - 1554	FP

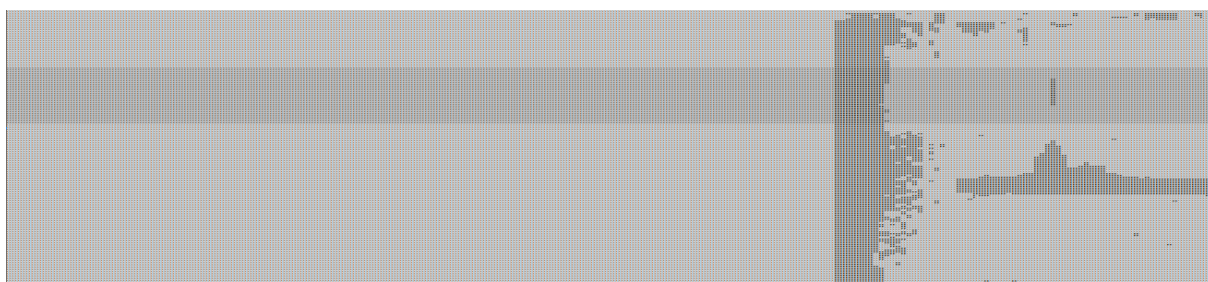


Figura 52: Resultado da análise do movimento de joelhos para o vídeo 1.3 (quadros 1 - 437).



Figura 53: Resultado da análise do movimento de joelhos para o vídeo 1.3 (quadros 438 - 874).

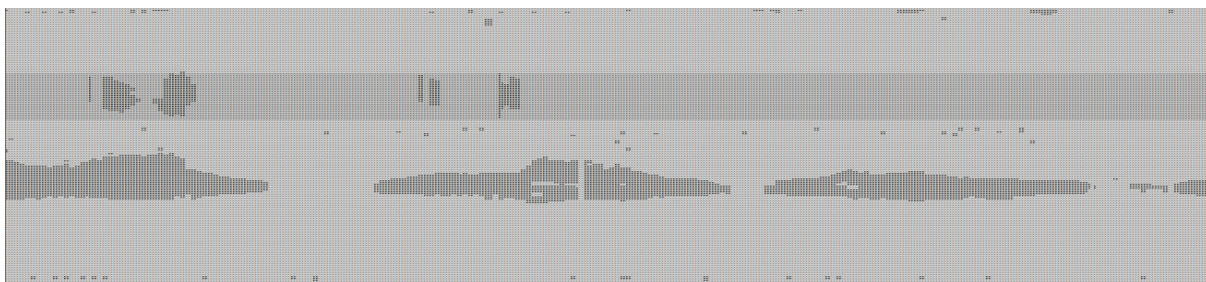


Figura 54: Resultado da análise do movimento de joelhos para o vídeo 1.3 (quadros 875 - 1311).

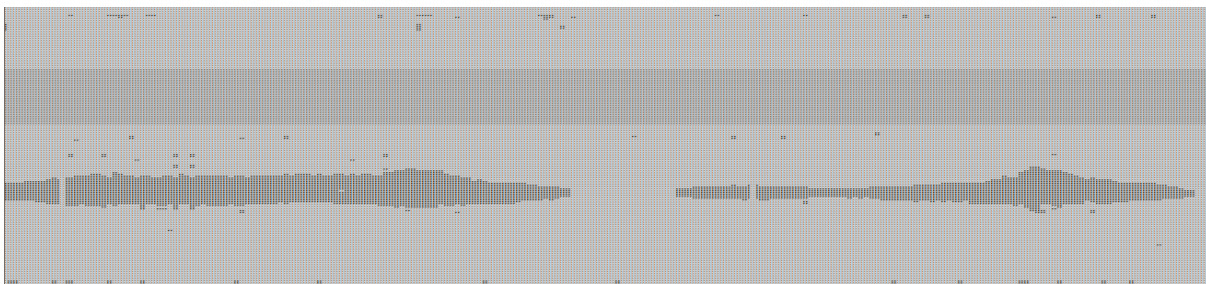


Figura 55: Resultado da análise do movimento de joelhos para o vídeo 1.3 (quadros 1749 - 2185).



Figura 56: Resultado da análise do movimento de joelhos para o vídeo 1.3 (quadros 2186 - 2622).



Figura 57: Resultado da análise do movimento de joelhos para o vídeo 2.1 (quadros 1 - 395).



Figura 58: Resultado da análise do movimento de joelhos para o vídeo 2.1 (quadros 396 - 790).



Figura 59: Resultado da análise do movimento de joelhos para o vídeo 2.1 (quadros 791 - 1185).

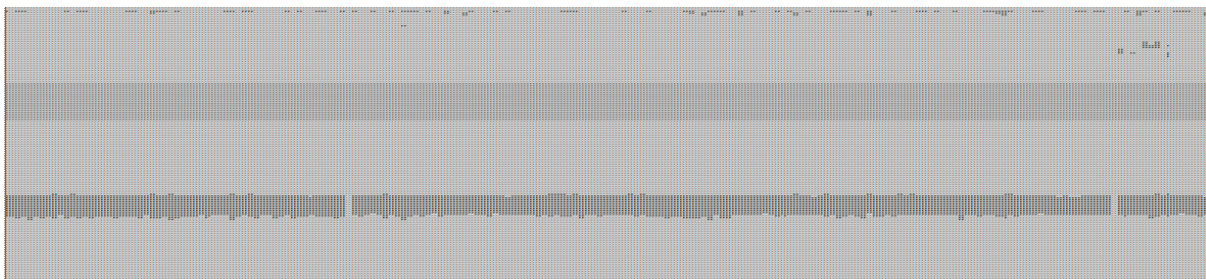


Figura 60: Resultado da análise do movimento de joelhos para o vídeo 2.1 (quadros 1186 - 1580).

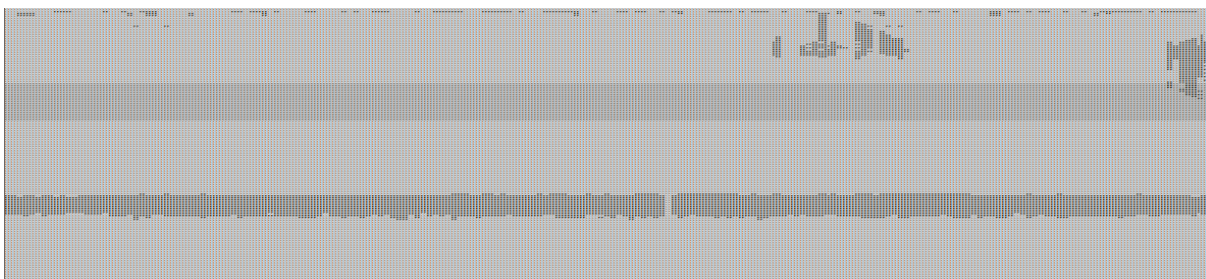


Figura 61: Resultado da análise do movimento de joelhos para o vídeo 2.1 (quadros 1581 - 1975).

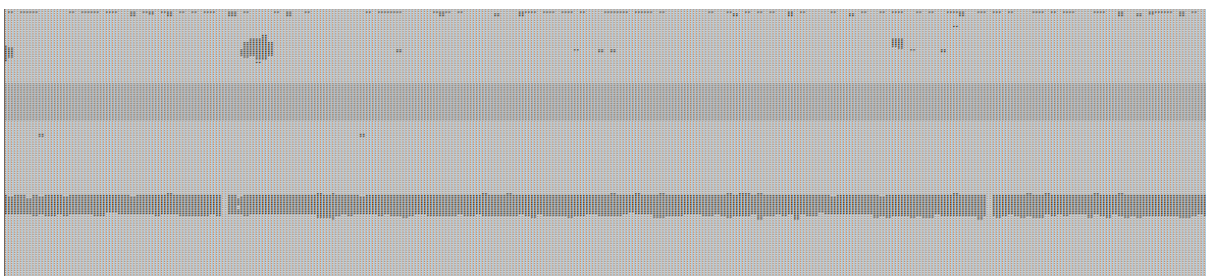


Figura 62: Resultado da análise do movimento de joelhos para o vídeo 2.1 (quadros 1976 - 2370).

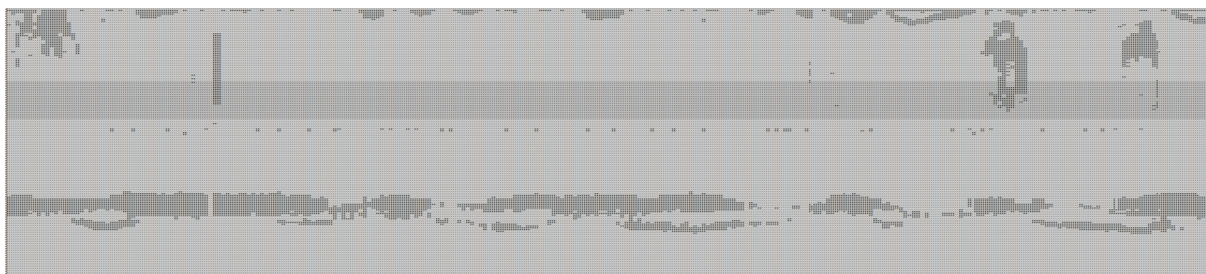


Figura 63: Resultado da análise do movimento de joelhos para o vídeo 2.2 (quadros 510 - 1072).

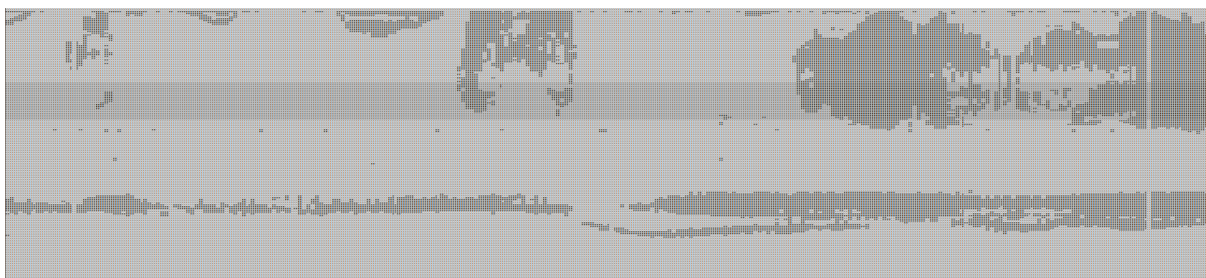


Figura 64: Resultado da análise do movimento de joelhos para o vídeo 2.2 (quadros 1073 - 1634).

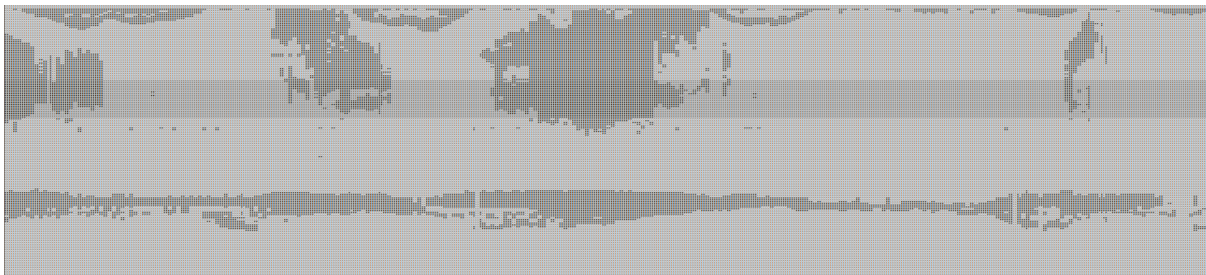


Figura 65: Resultado da análise do movimento de joelhos para o vídeo 2.2 (quadros 1635 - 2196).

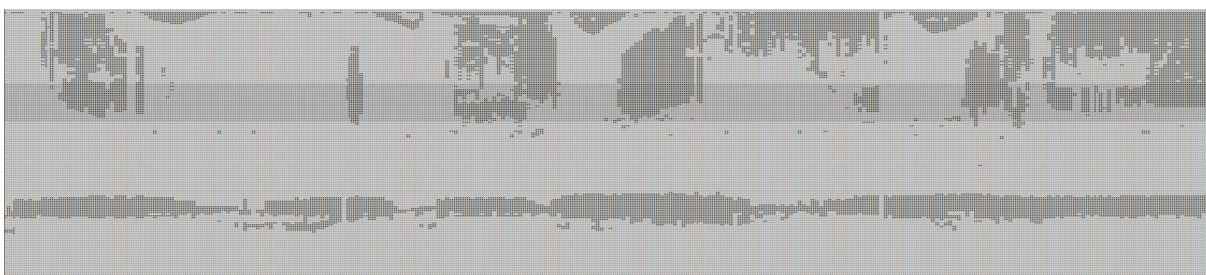


Figura 66: Resultado da análise do movimento de joelhos para o vídeo 2.2 (quadros 2197 - 2758).



Figura 67: Resultado da análise do movimento de joelhos para o vídeo 2.3 (quadros 210 - 646).

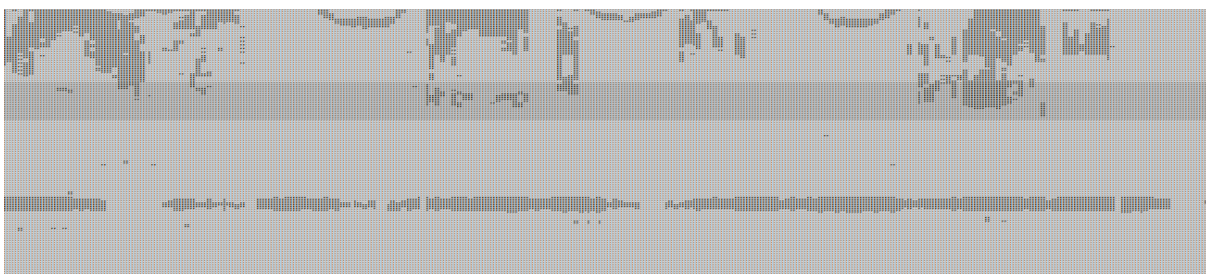


Figura 68: Resultado da análise do movimento de joelhos para o vídeo 2.3 (quadros 647 - 1082).

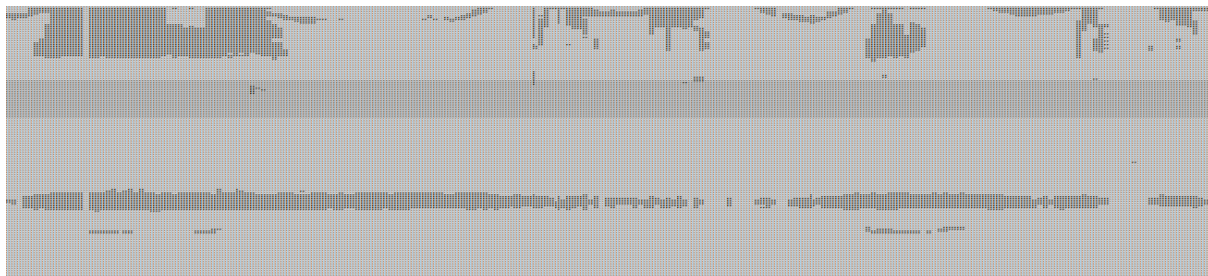


Figura 69: Resultado da análise do movimento de joelhos para o vídeo 2.3 (quadros 1083 - 1518).

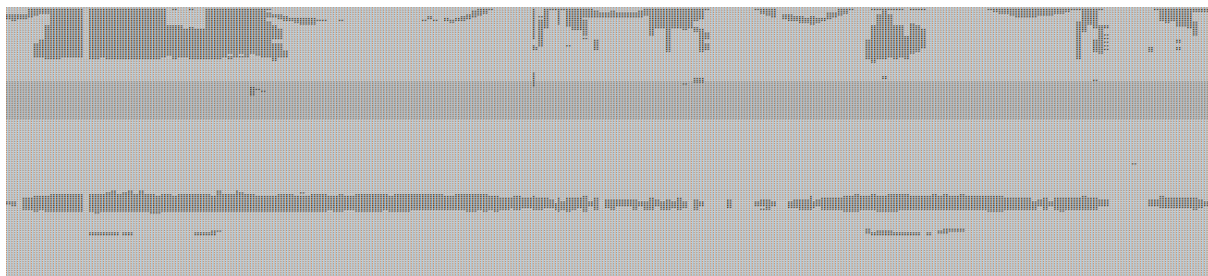


Figura 70: Resultado da análise do movimento de joelhos para o vídeo 2.3 (quadros 1519 - 1954).

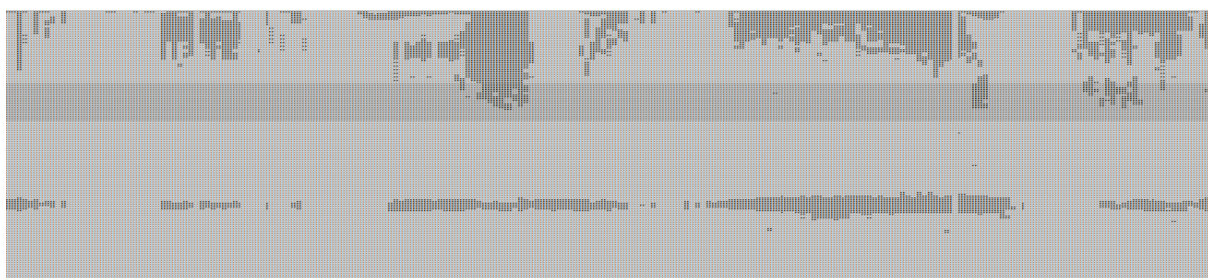


Figura 71: Resultado da análise do movimento de joelhos para o vídeo 2.3 (quadros 1955 - 2392).

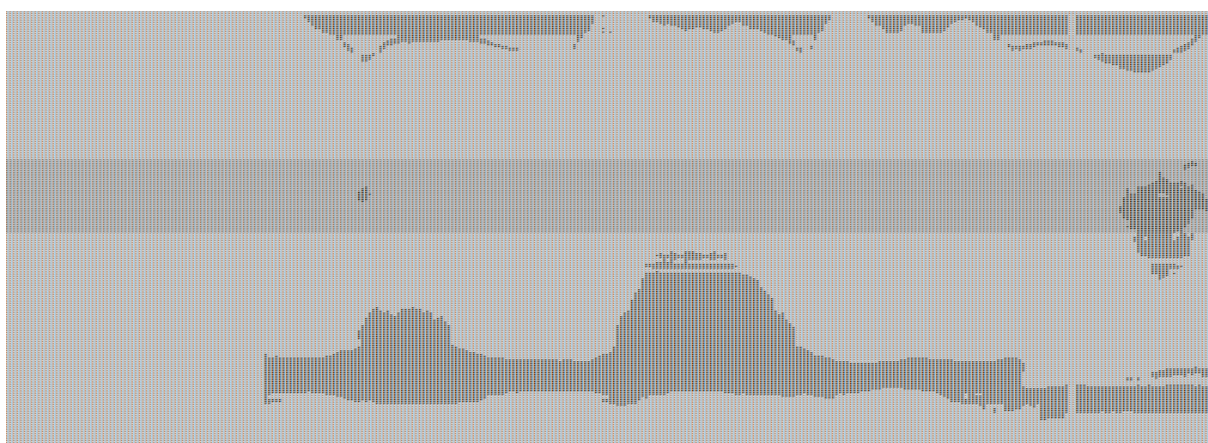


Figura 72: Resultado da análise do movimento de joelhos para o vídeo 3 (quadros 1 - 337).



Figura 73: Resultado da análise do movimento de joelhos para o vídeo 3 (quadros 338 - 674).

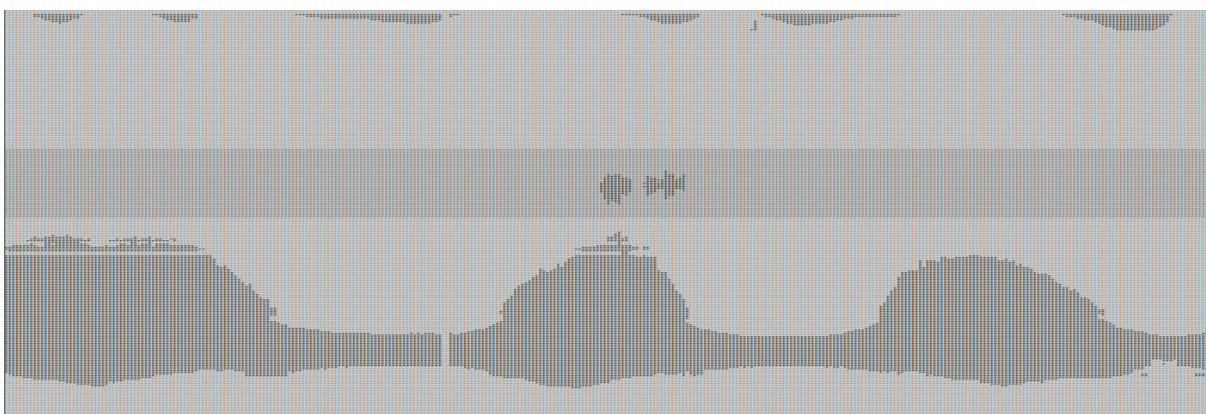


Figura 74: Resultado da análise do movimento de joelhos para o vídeo 3 (quadros 675 - 1011).

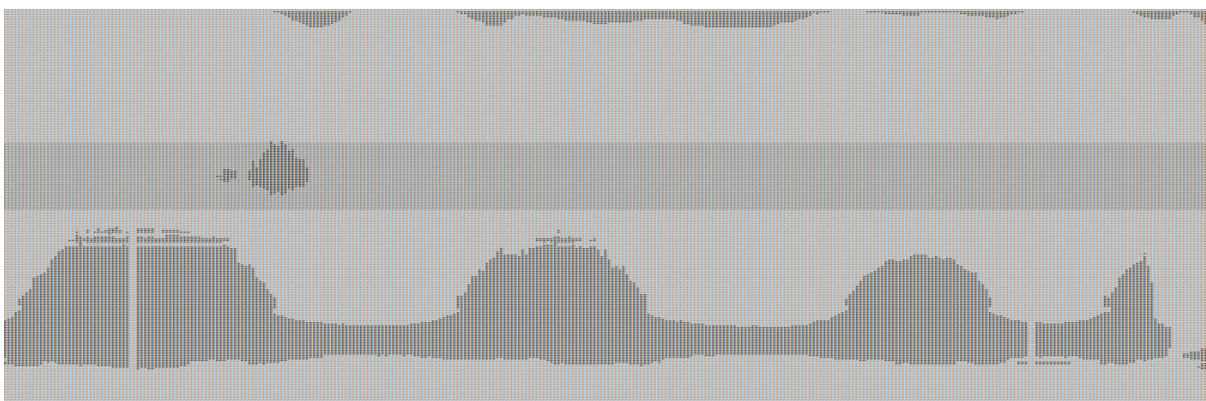


Figura 75: Resultado da análise do movimento de joelhos para o vídeo 3 (quadros 1012 - 1348).

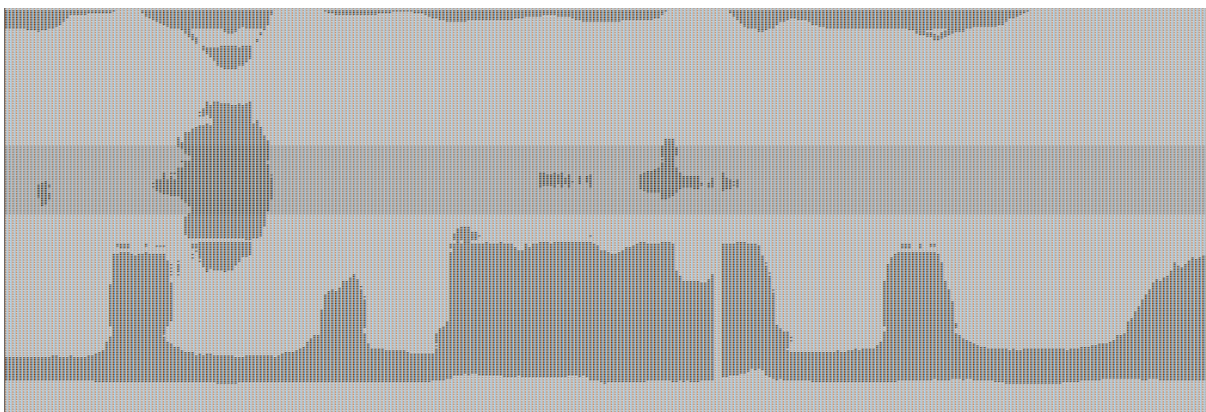


Figura 76: Resultado da análise do movimento de joelhos para o vídeo 3 (quadros 1349 - 1685).

Tabela 25: Resultado da análise do movimento de ondulação das costas para o vídeo 1.1.

GESTO	Intervalo de Quadros		RESULTADO
	ROTULAÇÃO	ALGORITMO	
1	159 - 167	170 - 177	PV
2	218 - 227	228 - 243	PV
3	272 - 281	290 - 296	PV
4	336 - 342	328 - 351	PV
5	383 - 390		FN
6		478 - 519	PV
7	543 - 550		FN
8	610 - 618	608 - 633	PV
9	662 - 670	674 - 693	PV
10	720 - 731		FN
11	796 - 805	800 - 817	PV
12	848 - 854	856 - 861	PV
13	904 - 912		FN
14		1053 - 1078	PV
15	1108 - 1116	1117 - 1132	PV
16	1171 - 1180	1189 - 1190	PV
17	1231 - 1240	1237 - 1262	PV
18	1291 - 1301	1289 - 1300	PV
19	1346 - 1355	1355 - 1370	PV
20	1404 - 1412	1421 - 1432	PV
21	1509 - 1513	1499 - 1502	PV
22	1559 - 1563	1563 - 1590	PV
23	1614 - 1624	1615 - 1632	PV
24	1670 - 1678	1671 - 1692	PV
25	1718 - 1726	1733 - 1748	PV
26	1781 - 1790	1779 - 1796	PV
27	1837 - 1845	1845 - 1852	PV
28	1893 - 1901	1909 - 1920	PV
29	1952 - 1964	1949 - 1976	PV
30	2001 - 2012	2013 - 2016	PV
31	2061 - 2071	2064 - 2091	PV
32	2115 - 2126	2124 - 2163	PV

Tabela 26: Resultado da análise do movimento de ondulação das costas para o vídeo 1.2.

GESTO	Intervalo de Quadros		RESULTADO
	ROTULAÇÃO	ALGORITMO	
1		228 - 229	PV
2	276 - 283	256 - 281	PV
3	371 - 380	362 - 369	PV
4	417 - 425	406 - 431	PV
5		508 - 509	PV
6		552 - 593	PV
7		624 - 659	PV
8		694 - 709	PV
9		746 - 747	PV
10	773 - 784	760 - 783	PV
11	902 - 912	894 - 923	PV
12	1034 - 1039	1020 - 1039	PV
13	1089 - 1099	1081 - 1100	PV
14	1157 - 1164	1141 - 1172	PV
15		1217 - 1230	PV
16		1373 - 1382	PV
17	1500 - 1507	1479 - 1508	PV
18	1603 - 1613	1619 - 1622	PV
19	1705 - 1716	1695 - 1712	PV
20	1759 - 1775	1753 - 1780	PV
21		1795 - 1802	PV
22		1819 - 1844	PV
23		1871 - 1874	FP
24		1905 - 1922	FP
25		1961 - 1968	PV
26		1997 - 2022	PV
27		2062 - 2065	PV

Tabela 27: Resultado da análise do movimento de ondulação das costas para o vídeo 1.3.

GESTO	Intervalo de Quadros		RESULTADO
	ROTULAÇÃO	ALGORITMO	
1		668 - 683	PV
2		728 - 739	PV
3	785 - 789	780 - 791	PV
4	836 - 840	822 - 831	PV
5	874 - 880	860 - 879	PV
6	938 - 942	916 - 941	PV
7	1019 - 1022	976 - 1013	PV
8		1048 - 1050	PV
9		1081 - 1094	PV
10	1156 - 1160	1153 - 1160	PV
11	1205 - 1214	1195 - 1212	PV
12	1253 - 1262	1241 - 1264	PV
13		1279 - 1280	FP
14	1300 - 1306	1303 - 1312	PV
15	1356 - 1360	1343 - 1362	PV
16		1375 - 1404	FP
17	1474 - 1478	1441 - 1472	PV
18	1517 - 1520	1507 - 1514	PV
19	1579 - 1583		FN
20	1654 - 1659	1627 - 1646	PV
21	1691 - 1695	1663 - 1684	PV
22		1721 - 1736	PV
23		1783 - 1792	PV
24	1868 - 1871	1849 - 1866	PV
25	1981 - 1986	1967 - 1988	PV
26	2022 - 2029	2013 - 2028	PV
27	2081 - 2092		FN
28	2145 - 2150	2144 - 2149	PV
29	2194 - 2201	2184 - 2209	PV
30	2253 - 2265	2248 - 2277	PV
31	2318 - 2335	2312 - 2331	PV
32	2384 - 2388	2366 - 2383	PV
33	2434 - 2440	2412 - 2439	PV
34	2500 - 2510	2484 - 2491	PV
35	2557 - 2563	2542 - 2567	PV

Tabela 28: Resultado da análise do movimento de ondulação das costas para o vídeo 1.4.

GESTO	Intervalo de Quadros		RESULTADO
	ROTULAÇÃO	ALGORITMO	
1	156 - 159		FN
2	203 - 209	204 - 208	PV
3	241 - 246	245 - 248	PV
4	273 - 277		FN
5	324 - 329	325 - 333	PV
6	382 - 387	383 - 387	PV
7	430 - 434	431	PV
8	465 - 468	453 - 470	PV
9	500 - 507		FN
10	539 - 545	541 - 545	PV
11	571 - 577	567 - 581	PV
12	608 - 616	613 - 620	PV
13	661 - 668	660 - 690	PV
14	756 - 763	748 - 753	PV
15	797 - 805	805	PV
16	846 - 853		FN
17	905 - 911	883 - 914	PV
18	941 - 948		FN
19	995 - 1000	985 - 1007	PV
20	1035 - 1045	1027 - 1037	PV
21	1090 - 1095	1089	PV
22	1173 - 1179		FN
23	1230 - 1234		FN
24	1268 - 1272	1257 - 1281	PV
25	1320 - 1324		FN
26	1362 - 1369		FN
27	1409 - 1417	1407 - 1428	PV
28	1444 - 1452	1448 - 1454	PV
29	1499 - 1509		FN
30	1559 - 1567	1549 - 1574	PV
31	1610 - 1617		FN
32	1657 - 1661	1657	PV
33		1684 - 1702	PV

Tabela 29: Resultado da análise do movimento de ondulação das costas para o vídeo 2.3.

GESTO	Intervalo de Quadros		RESULTADO
	ROTULAÇÃO	ALGORITMO	
1		340 - 349	PV
2		412 - 413	PV
3		422 - 435	FP
4		516 - 537	PV
5	606 - 611	614 - 615	PV
6	769 - 775	762 - 783	PV
7	875 - 881	858 - 885	PV
8	950 - 955	936 - 967	PV
9		1061 - 1062	FP
10	1122 - 1131	1102 - 1160	PV
11	1351 - 1356		FN
12	1502 - 1507	1505 - 1522	PV
13	1645 - 1650	1621 - 1670	PV
14	1809 - 1814	1787 - 1826	PV

Tabela 30: Resultado da análise do movimento de ondulação das costas para o vídeo 3.

GESTO	Intervalo de Quadros		RESULTADO
	ROTULAÇÃO	ALGORITMO	
1	349 - 363	338 - 365	PV
2		475 - 495	PV
3		746 - 808	FP
4	1073 - 1085	1043 - 1090	PV
5	1286 - 1297		FN
6		1630 - 1647	PV
7	1871 - 1892	1860 - 1901	PV

Tabela 31: Resultado da análise do movimento de ondulação das costas para o vídeo 4.

GESTO	Intervalo de Quadros		RESULTADO
	ROTULAÇÃO	ALGORITMO	
1	103 - 108	91 - 124	PV
2	164 - 167	128 - 161	PV
3	229 - 233	214 - 230	PV
4	285 - 288	276 - 325	PV
5	412 - 415	397 - 433	PV
6	509 - 512	497 - 507	PV
7	542 - 546	532 - 546	PV
8	579 - 582	570 - 606	PV
9		643 - 653	FP
10		901 - 907	FP
11	990 - 995	980 - 1004	PV
12	1103 - 1108	1092 - 1110	PV
13		1215 - 1216	FP
14	1523 - 1526	1513 - 1529	PV
15	1621 - 1624	1568 - 1639	PV

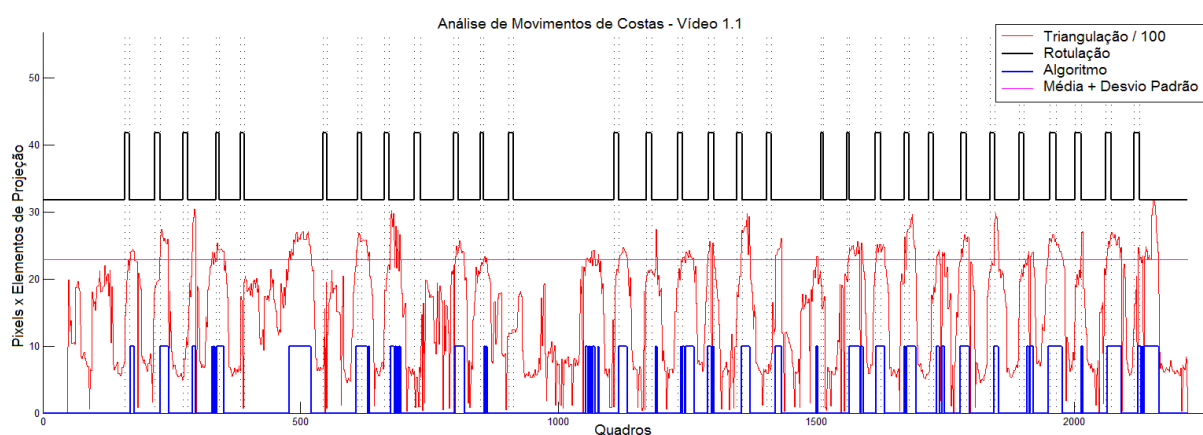


Figura 77: Resultado da análise do movimento de ondulação das costas para o vídeo 1.1.

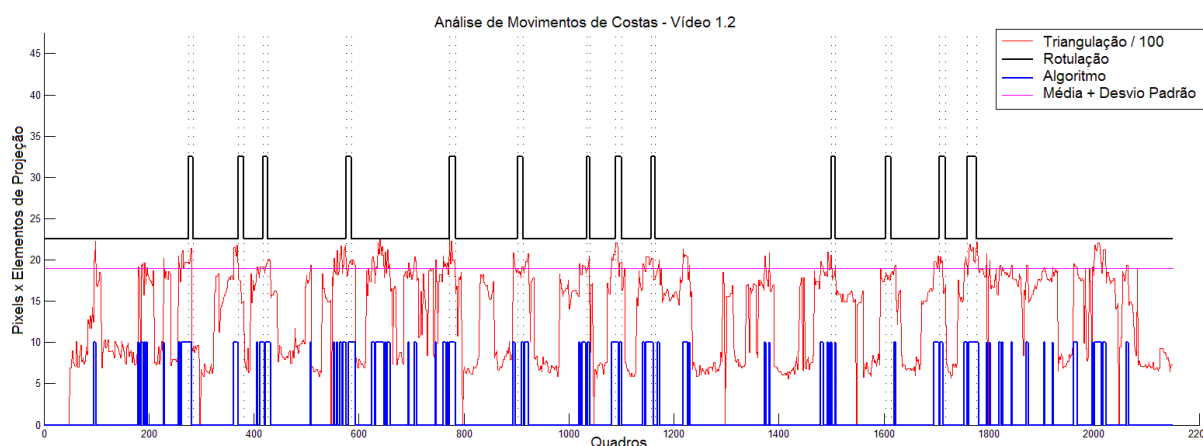


Figura 78: Resultado da análise do movimento de ondulação das costas para o vídeo 1.2.

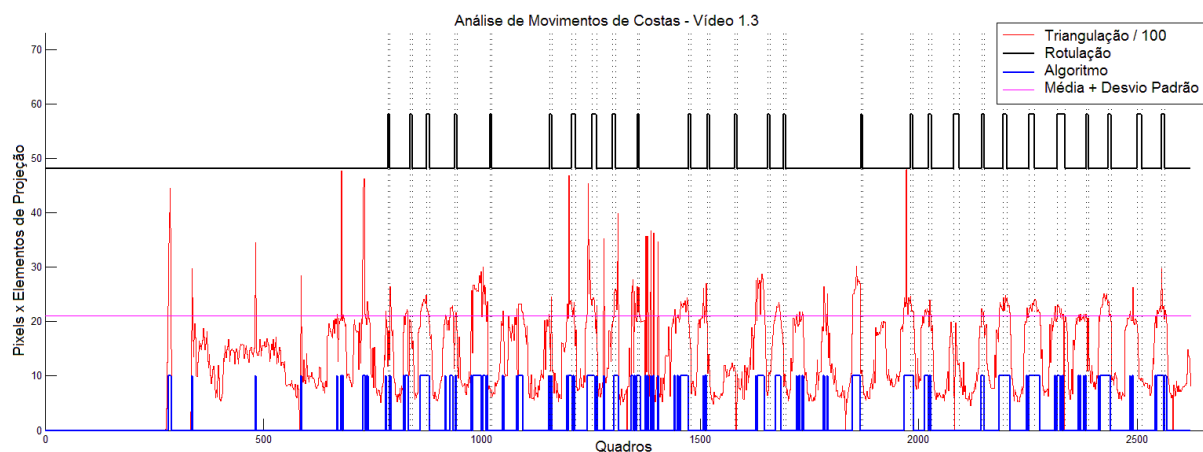


Figura 79: Resultado da análise do movimento de ondulação das costas para o vídeo 1.3.

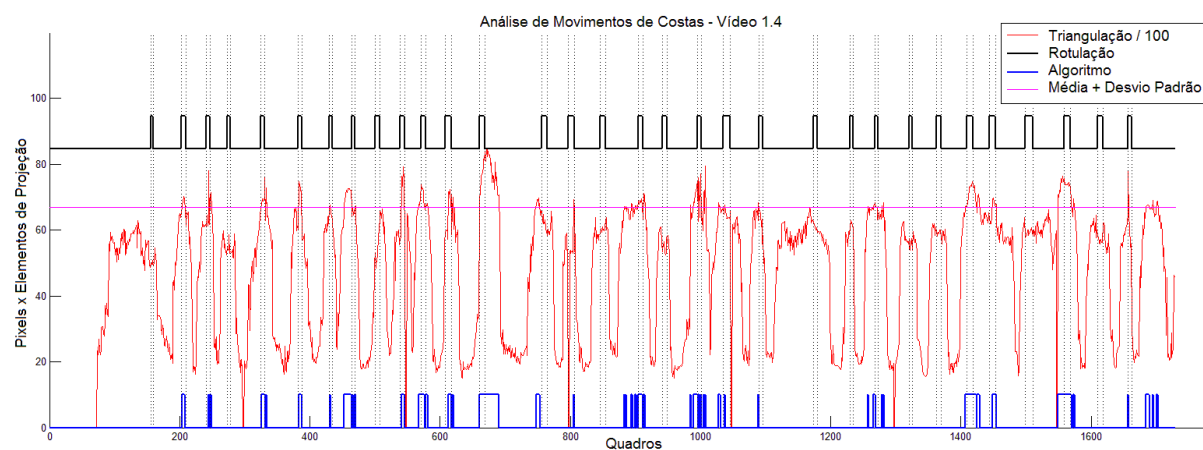


Figura 80: Resultado da análise do movimento de ondulação das costas para o vídeo 1.4.

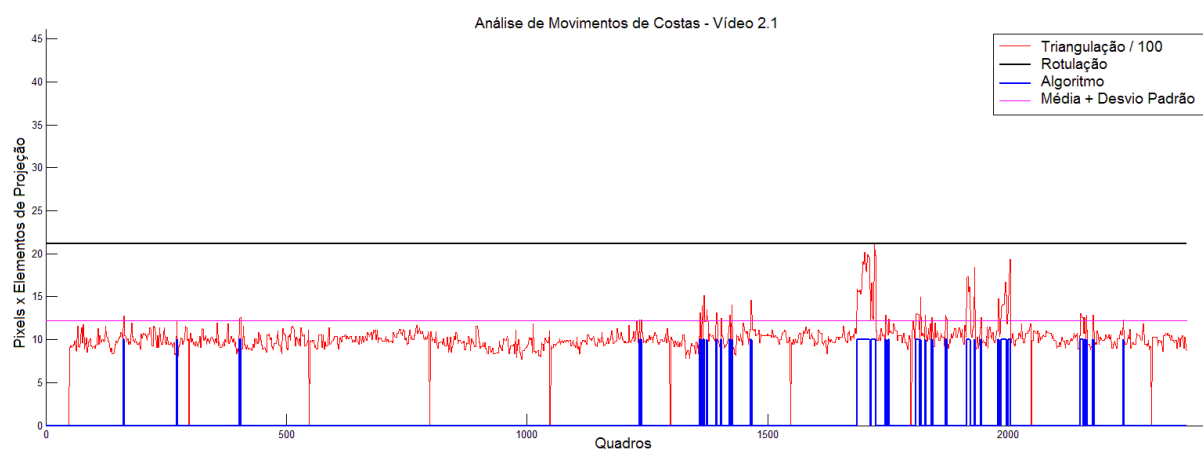


Figura 81: Resultado da análise do movimento de ondulação das costas para o vídeo 2.1.

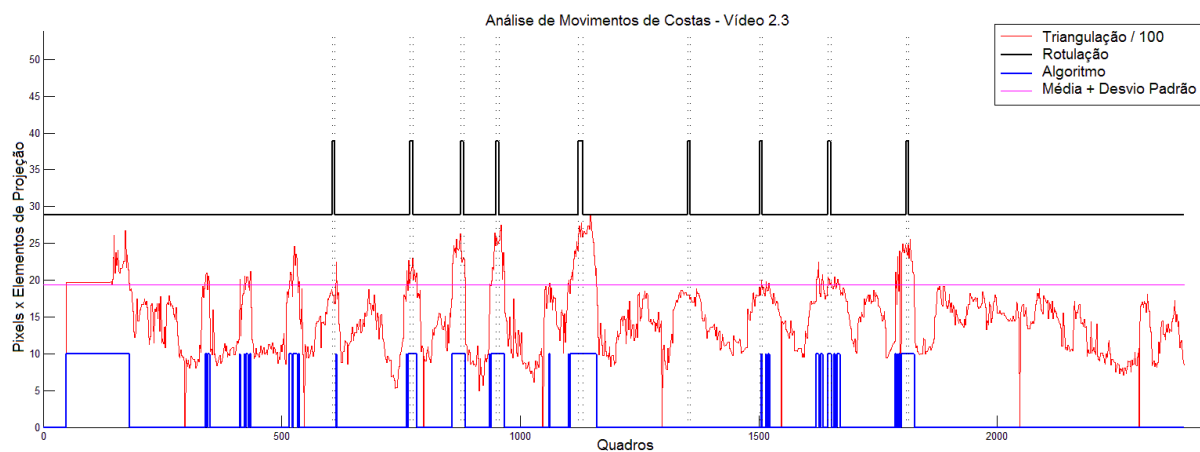


Figura 82: Resultado da análise do movimento de ondulação das costas para o vídeo 2.3.

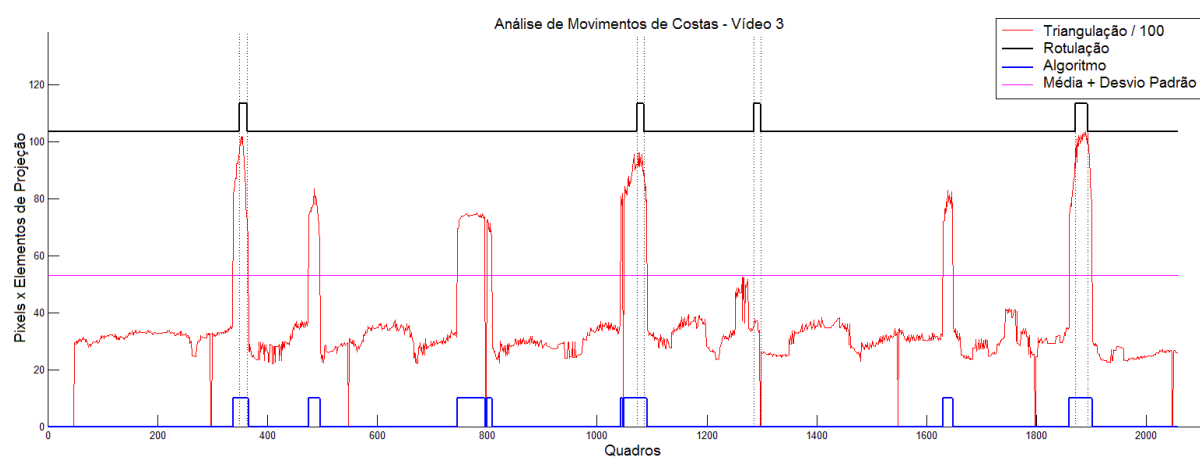


Figura 83: Resultado da análise do movimento de ondulação das costas para o vídeo 3.

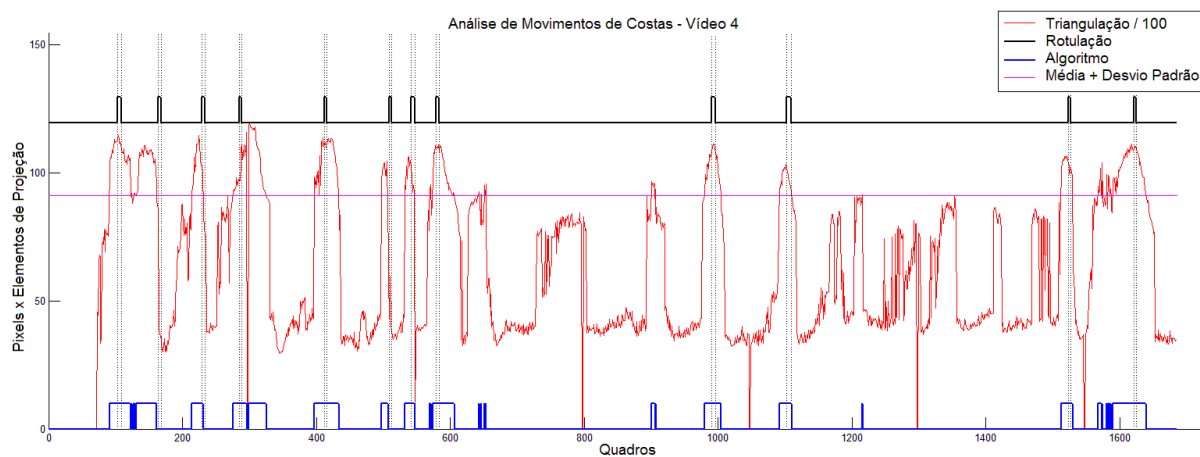


Figura 84: Resultado da análise do movimento de ondulação das costas para o vídeo 4.